

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 05:29:18 ; Search time 76 Seconds
(without alignments)
5399.125 Million cell updates/sec

Title: US-09-806-536A-29

Perfect score: 1338

Sequence: 1 ggtgagcgcagctgtccga.....ggggcagcagcatatggggt 1338

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

- 1: /cgn2.6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2.6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2.6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2.6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq:*
- 6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201.8	15.1	1742	3	US-09-099-676-2
2	201.8	15.1	1742	4	US-09-565-910-2
3	75.8	5.7	28171	4	US-08-961-527-22
4	73.4	5.5	4403765	4	US-09-103-840A-2
5	73.4	5.5	4411529	4	US-09-103-840A-1
6	66.6	5.0	1147	3	US-08-665-716-1
7	42.8	3.2	466	3	US-09-154-083-17
8	42.2	3.2	7218	1	US-08-232-463-14
9	41.2	3.1	1926	4	US-09-249-585A-2
10	41.2	3.1	2580	3	US-09-050-863-2
11	41.2	3.1	2580	4	US-09-359-081-2
12	41.2	3.1	5452	2	US-09-130-114-1
13	41.2	3.1	9600	4	US-08-910-647-1
14	41.2	3.1	9600	4	US-09-620-925-1
15	41.2	3.1	10596	1	US-07-884-811-15
16	41.2	3.1	10596	1	US-07-885-971-15
17	41.2	3.1	10596	1	US-08-087-783A-15
18	41.2	3.1	10596	1	US-08-194-088B-15
19	41.2	3.1	10596	2	US-08-194-087-15
20	41.2	3.1	10596	5	PCT-US93-04548-15
21	40.8	3.0	3546	1	US-07-951-715A-14
22	40.8	3.0	3546	2	US-08-459-448A-14
23	40.8	3.0	3546	3	US-08-459-504B-14
24	40.8	3.0	3546	3	US-08-459-504B-14
25	40.8	3.0	3546	3	US-08-459-444-14
26	40.8	3.0	3546	4	US-09-547-422-14
27	40.6	3.0	1820	1	US-08-173-508-7

Sequence 7, Appli
Sequence 7, Appli
Patent No. 5352575
Sequence 1, Appli
Sequence 5, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 12, Appli

28 40.6 3.0 1821 2 US-08-265-310-7
29 40.6 3.0 1821 3 US-08-951-742-7
30 39.6 3.0 1053 6 5352575-6
31 39.6 3.0 1146 1 US-08-482-385A-1
32 39.6 3.0 2728 1 US-08-482-385A-5
33 39.4 2.9 2127 1 US-08-464-340A-1
34 39.4 2.9 2127 5 PCT-US94-08449A-1
35 39.2 2.9 3468 1 US-07-951-715A-2
36 39.2 2.9 3468 2 US-08-459-448A-2
37 39.2 2.9 3468 3 US-08-459-504B-2
38 39.2 2.9 3468 3 US-08-459-504B-2
39 39.2 2.9 3468 3 US-09-053-549-3
40 39.2 2.9 3468 4 US-09-547-422-2
41 39.2 2.9 3546 1 US-07-951-715A-12
42 39.2 2.9 3546 2 US-08-459-448A-12
43 39.2 2.9 3546 3 US-08-459-504B-12
44 39.2 2.9 3546 3 US-08-459-504B-12
45 39.2 2.9 3546 3 US-08-459-504B-12

ALIGNMENTS

RESULT 1

US-09-099-676-2
; Sequence 2, Application US/09099676
; Patent No. 6100075
; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Corley, Neil C.

; APPLICANT: Baughn, Mariah R.

; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE

; TITLE OF INVENTION: HOMOLOG

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/099,676

; FILING DATE: HERewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Cerrone, Michael C

; REGISTRATION NUMBER: 39,132

; REFERENCE/DOCKET NUMBER: PF-0532 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-855-0572

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1742 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: PROSNO1

; CLONE: 2278458

US-09-099-676-2

Query Match 15.1%; Score 201.8; DB 3; Length 1742;

Best Local Similarity 55.0%; Pred. No. 1.2e-38;
Matches 448; Conservative 0; Mismatches 352; Indels 15; Gaps 2;

QY 42 GAGCGCTTCCCGCGCGTGGGCTTCTGTCGGCGCGGCGGCATGCGGGGGCCATCGCG 101
DB 180 GAACCGCGGACCATGAGCGTGGGCTTATCGGGCGCGGCGAGCTGATGCTGCTGGCG 239

QY 102 CAGGCGCTATCAGAGCAGGAAAGTGAAGCTCAGCACATCTAGGCGAGTGCCACACA 161
DB 240 CGGGGCTTACGCGCGGAGGATCTGTCGGCTCAAGATATAGCAGCTGCCAGAA 299

QY 162 GACAGGAACTTATGTCACCTTCAAGCTTGGGTTGCGGACCACTCCCAAC 215
DB 300 ATGAACCTGCCACCGGTGTCGGCTCAGGAAGTGGGTGAACCTGACACGCGAGCAAC 359

QY 216 CAGGAGGTGCTCAGAGTGCCTGCTGCTATCTTTGCCCAACAGCCTCATGTCGTGCA 275
DB 360 AAGGAGCGGTGAAGCAGCGACGCTCTGTTCTGCTGTGAAGCCACATATATCCCC 419

QY 276 GCTGCTGCGGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 335
DB 420 TTATCTGCTGATGAGTGGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479

QY 336 GCTGGGCTGCTCAGACACCTTGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 386
DB 480 GCTGGTGTACCATCAGCTCTGTCGAGAGAAAGCTGATGCGATTCACGCGAGCCCAAA 539

QY 387 GTGCTGCGGGTCTTGGCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446
DB 540 GTGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599

QY 447 GCGGGCGCCACGTCGGGAGCAGGAGCAAGCTCTCTGAGCATCTGCTGAGGCTGT 506
DB 600 ACGGCGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659

QY 507 GGGCGGTGTGAGGAGTGTGCTGAAGCTACGTCGACATCCACACTGGCTCAGTGGCAGT 566
DB 660 GCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719

QY 567 GCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 626
DB 720 GGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779

QY 627 ATGCCACGAGCTGGCCCGCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 686
DB 780 TTGCCACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839

QY 687 CTGCTGACGAGGCGCAACACCCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 746
DB 840 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 899

QY 747 ACCACCATGATGAGTCTCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
DB 900 GGCACCATCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959

QY 807 GCGTGGAGTGGCCACCTGCGCGGCGCAAGAGCT 841
DB 960 GCAGTTGAGGCTCTCTGTATCCGAAACAGAGAGCT 994

RESULT 2
US-09-565-910-2
; Sequence 2, Application US/09565910
; Patent No. 6268192
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/565,910
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/099,676
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Cerrone, Michael C
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0532 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-855-0572
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1742 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNOM01
CLONE: 2278458
US-09-565-910-2

Query Match 15.1%; Score 201.8; DB 4; Length 1742;
Best Local Similarity 55.0%; Pred. No. 1.2e-38;
Matches 448; Conservative 0; Mismatches 352; Indels 15; Gaps 2;

QY 42 GAGCGCTTCCCGCGCGTGGGCTTCTGTCGGCGCGGCGGCATGCGGGGGCCATCGCG 101
DB 180 GAACCGCGGACCATGAGCGTGGGCTTATCGGGCGCGGCGAGCTGATGCTGCTGGCG 239

QY 102 CAGGCGCTATCAGAGCAGGAAAGTGAAGCTCAGCACATCTAGGCGAGTGCCACACA 161
DB 240 CGGGGCTTACGCGCGGAGGATCTGTCGGCTCAAGATATAGCAGCTGCCAGAA 299

QY 162 GACAGGAACTTATGTCACCTTCAAGCTTGGGTTGCGGACCACTCCCAAC 215
DB 300 ATGAACCTGCCACCGGTGTCGGCTCAGGAAGTGGGTGAACCTGACACGCGAGCAAC 359

QY 216 CAGGAGGTGCTCAGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 275
DB 360 AAGGAGCGGTGAAGCAGCGACGCTCTGTTCTGCTGTGAAGCCACATATATCCCC 419

QY 276 GCTGCTGCGGAGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 335
DB 420 TTATCTGCTGATGAGTGGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479

QY 336 GCTGGGCTGCTCAGACACCTTGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 386
DB 480 GCTGGTGTACCATCAGCTCTGTCGAGAGAAAGCTGATGCGATTCACGCGAGCCCAAA 539

QY 387 GTGCTGCGGGTCTTGGCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446
DB 540 GTGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599

QY 447 GCGGGCGCCACGTCGGGAGCAGGAGCAAGCTCTCTGAGCATCTGCTGAGGCTGT 506
DB 600 ACGGCGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659

QY 507 GGGCGGTGTGAGGAGTGTGCTGAAGCTACGTCGACATCCACACTGGCTCAGTGGCAGT 566

Db 660 GCCTTCGACCTGAGGTGGAAGAGGACCTCATCGATGCGCTCAGGGGCTCAGTGGCAGC 719
QY 567 GCGTGGGCTTCGTGTGTGCAATTCCTCCAGAGCCCTGGTGAAGAGCGCTCAAGATGGGC 626
Db 720 GGGCTCGCTATGCAATTCATGCTCTGACGCAATGGCTGATGGTGGGTGAAGATGGGT 779
QY 627 ATGCCAGCAGCTGGCCACCGCATCGCTGCCAGACCCCTGCTGGGACCGCCAGATG 686
Db 780 TTGCCACGGCGCTGGCAATCAACTCGGGGCCAGGCTTTGCTGGGAGCTGCCAAGATG 839
QY 687 CTGCTGCACGAGGGCCAAACACCCAGCCAGCTGCGCTCAGACGCTGTGCACCCCGGTGGC 746
Db 840 CTGCTGGACTCGGACGACATCCATGCGCAGCTTAAGGACATGTCTGCTCCCTGGGGGA 899
QY 747 ACCACATCTATGACTCTCAGCCCTGGAGCAGGGCGGGCTGCGAGCAGCCACCATGAGC 806
Db 900 GCCACATCCAGCCCTGACCTTTCTAGAGAGTGGGGCTTCCGCTCTCTGCTCATCAAT 959
QY 807 GCGCTGGAGGCTGCCACCTGCCGGGCCAAGAGCT 841
Db 960 CGAGTTGAGGCTCTCTGTATCGGAACAGCAGAGCT 994

RESULT 3

US-08-961-527-22
; Sequence 22, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 28171 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-22

Query Match

Best Local Similarity 5.78; Score 75.8; DB 4; Length 28171;

Matches 260; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

QY 319 TCTTGGTGTGCGTGGCTGGGTGCTCTGTGAGCAGCCCTGGAGGAGCTCTGCGCCCA 378
Db 7365 TTTTGATTGATGGCAGCTGGATTGACCTTAGAAAACACTAGCAAGTCTTATCCCAAGTC 7424

QY 379 ACACACGGGTGCTGCGGGTCTTGCCCAACCTGCCCCTGTGTGGTCCAGAAAGGCCCATAG 438
Db 7425 AACACCGAAATATTTCGTATGATGCTTAATACCCCTGCTTCTATCGGCAAGGATGATTA 7484
QY 439 TGATGGCGGGGGCCGACGCTGGGAGCAGGACCAAGCTCTCTGACGATCTGCTGG 498
Db 7485 GTTATGCTTGTCTCTTAATTCAGGGCTGAGGACAGTGTGCTTTTATCAGCTTTTAG 7544
QY 499 AGGCTCTGGCGGTGTGAGGAGTGTCTGAAGCTACGTCGACATCCACACTGGCTCA 558
Db 7545 CCRAAGGTGTCTCTTGGTTGAACCTAGGAGAAATTAATCGATGAGCCAGGCTTG 7604
QY 559 GTGCACTGGCGGTGCGCTCTGCTGTGCTATCTCCGAGGCGCTGGCTGAAGAGCGCTCA 618
Db 7605 CAGCTTGTGACACGCTTTTGTCTATCTTTTATCGAGGCTTGGCAGATGCAAGTGTTC 7664
QY 619 AGATGGCATGCCACGAGCTGGCCACCGCATCGCTGCCAGACCCCTGCTGGGACGG 678
Db 7665 AGACAGGATTACCACGAGAAATAGCATTTGAAATGGCAGCACAACCTGTGTAGGAGCTG 7724
QY 679 CCAAGATGCTGCTGCACGAGGGCCACACACCCAGCCAGCTGCGCTCAGACGCTGTCACCC 738
Db 7725 GSCAATTGGTCTTGAAAGTCAGCAACATCTCTGGAGTATTGAAGACCAAGTCTGTAGCC 7784
QY 739 CGGTGGCACACCATCTATGAGCTCCACCCCTGGAGCAGGCGGGCTCGGAGCAGCCA 798
Db 7785 CAGCGGTTCGACTATGCTGTGTAGCAAGCCTAGAAAGCCATGCTTTCCGAGGAACAG 7844
QY 799 CCATGAGCGCGTGGAGGCTGCCACCTGCCGCGCAAGGAGCTCAGCAGAAAGTAGGCTG 858
Db 7845 TCAATGATGAGTTCATCAAGCCTCAACAGCAACAACTAGGTAATAAGAGGTAG 7904
QY 859 GGCTGTGGCATCTCTTCTGCTGCTG 885
Db 7905 TTTTGTGCTGCTCTTTTATGCTGGCTG 7931

RESULT 4

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 5.58; Score 73.4; DB 4; Length 4403765;
Best Local Similarity 48.78; Pred. No. 1.6e-07;
Matches 200; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY 284 GCAGAGGTGGCTCTCTGCTGCTCACCACCTCAACACATCTTGTGCTCGGTGCTGGGT 343
Db 591703 GACTCGGCGGGCGGAAACAGCAGTGTCTGAGCAGGTGTCTGTCACCGTGTAGCGGGCAT 591762
QY 344 GTCTCTGACACCCCTGGAGGAGTGTGCTGCCCCCAACACACGCGGTGCTCGCGGTCTTGCC 403

Db 591763 CACGATCGCGTATTTCGAATCCAAAGCTACCGCGGGACGCCAGTGGTGGCTGCATGCC 591822
QY 404 CAACTGCGCTGTGTGTCAGGAAGGGCCATAGTATGCGCGGGGCCGCCACAGTGGG 463
Db 591823 GAACGCGGGGCGATTGGTGGAGCGGTTACAGCGTGGCCAAAGCGCGCTTTGTAC 591882
QY 464 GAGCAGCGAGACCAAGCTCCTGACGATCTGCTGAGGCGCTGTGGCGGTGTGAGGAGGT 523
Db 591883 CCCGCAACAGCTTGAGGAGGTCTCGGCCCTTGTTCACCGGTGCGCGCGTGTGACCGT 591942
QY 524 GCGTGAAGCCTACGTGCGACATCCACACTGGCGCTCAGTGGCAGTGGCGTGGCGCTTCGTGTG 583
Db 591943 TCCGGAATCGCAGTTGGAGCGGCTGACCGCGGTGTCGGGCTCGGCGCTTATTTCTT 592002
QY 584 TGCATTCTCCGAGGCGCTGGCTGAAGGAGCGCTCAAGATGGCATGGCCAGCAGCGCTGGC 643
Db 592003 TCTGCTGGTCGAGGCGCTGGTGGATGCGCGAGTGGCGGTGGCGTTGACCGCTCAGGTGGC 592062
QY 644 CCACCGCATCGTGGCGGACCGCTCTGGGGACGCCAAGATGCTGCTGCA 694
Db 592063 CACCGATCTCGCGCGCAGACAATGGTGGCTCAGCGCGATGCTGCTGGA 592133

RESULT 5

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 5.5%; Score 73.4; DB 4; Length 4411529;
Best Local Similarity 48.7%; Pred. No. 1.6e-07;
Matches 200; Conservative 0; Mismatches 211; Indels 0; Gaps 0;
QY 284 GGCAGAGGTGGCTTCCTGTGTGTCACCACTGAACACATCTTGTGTGTCGGTGGCTGGGGT 343
Db 590341 GACTCGCGCGCGGAAACACAGTCTCAGCAGGTGTTCGTACCGTGGTAGCGGCAT 590400
QY 344 GTCTGTAGCACCTGGAGAGGTGTGTCGCCCAACACACAGCGGTGCTGGGGTCTTGCC 403
Db 590401 CACGATCGCGTATTTCCGAATCCAAAGCTACCGCTGGGACGCCAGTGGTGGCGATGCC 590460
QY 404 CAACCTGCCCTGTGTGTCAGGAAGGGCCATAGTATGGCGCGCGCGCCACGCTGGG 463
Db 590461 GAACGCGGGCGATTGGTGGAGCGGGGTACAGCGCTGGCCAAAGCGCGCTTTGTAC 590520
QY 464 GAGCAGCGAGACCAAGCTCCTGAGCATCTGTGGAGGCGCTGTGGCGGTGTGAGAGGT 523
Db 590521 CCCGCAACAGCTTGAGAGGTCTCGGCCCTTGTTCAGCGCGGTGCGCGCGGTGCTGACCGT 590580
QY 524 GCCTGAAGCCTAGCTGCATCCACACTGGCCTCAGTGCAGTGGCGTGGCGCTTGTGTG 583
Db 590581 TCCGGAATCGCAGTTGGAGCGGTGACCGCGGTGTCCGGCTCGGGTCCGGCGCTTATTTCTT 590640
QY 584 TGCATTCTCGAGGCGCTGGCTGAAGGAGCGCGTCAAGATGGCGATGGCCAGCAGCGCTGGC 643

Db 590641 TCTGCTGGTCGAGGCGCTGGTGGATCCGAGTTCGGGGTGGGCTTACGCGCTCAGGTGGC 590700
QY 644 CCACCGCATCGCTGCCACACACCTGCTGGGGACGGCCAAAGATGCTGCTGCA 694
Db 590701 CACCGATCTCGCGCGCAGACAATGGTGGCTCAGCGCGATGCTGCTGGA 590751
RESULT 6
US-08-665-716-1
; Sequence 1, Application US/08665716
; Patent No. 5789222
; GENERAL INFORMATION:
; APPLICANT: KELLY, ROSEMARIE
; APPLICANT: REGISTER, ELIZABETH A
; APPLICANT: MASUREKAR, PRAKASH S
; TITLE OF INVENTION: P5C REDUCTASE GENE FROM ZALERION
; TITLE OF INVENTION: ARBORICOLA
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MERCK & CO., INC.
; STREET: 126 E. LINCOLN AVENUE
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: US
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/665,716
; APPLICATION NUMBER: US/08/665,716
; FILING DATE: 23-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: KORSER, ELLIOTT
; REGISTRATION NUMBER: 32,705
; REFERENCE/DOCKET NUMBER: 19453PV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-594-5493
; TELEFAX: 908-594-4720
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1147 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 47..960
US-08-665-716-1

Query Match 5.0%; Score 66.6; DB 1; Length 1147;
Best Local Similarity 52.3%; Pred. No. 6.6e-07;
Matches 172; Conservative 0; Mismatches 154; Indels 3; Gaps 1;
QY 520 AGTGGCTCTGAAGCGCTAGTGTGCACATCCACACTGGCGCTCAGTGGCAGTGGCGCTTCG 579
Db 651 AGCTCCCGCTGTCTACATGATGATGCCAGCAGCGGCTGTGTGGCAGTGGCGCGCTTTT 710
QY 580 TGTGTGATCTCCGAGGCGCTGCTGAAGAGCGCTCAAGATGGGATGCCAGCAGCC 639
Db 711 TTGCGCTCATACTGAGGCTGCGATTGATGGGCTGTGGCGATGGGTGCGCAAGACAG 770
QY 640 TGGCCCGCAGCGATCGCTGCCAGACCGCTGCTGGGAGCGGCGCAAGATGCTGTCACGAGG 699
Db 771 AGGTCAGAGATGGCGCGCAGACTATGAAGGCGGTGCTGGATTGTTGTTCTGGAG 830
QY 700 GCCAACACCCAGCGCAGCTGGCTTCAGACGTGTGCACCGCGGTGGCAGCACCATCTATG 759


```
Db 831 ---AGCATCGCGGTTGCTGAAAGATAAAGTGACTACGCGCGGTGGTCCACGATAGGTG 887
QY 760 GACTCCACGCCCTGGAGCAGGCGGCTCGGAGCAGCCACCATGAGCGCGCTGGAGGTG 819
Db 888 GGTGATGTTGCTGGAAGAGGAGGCTGAGAGGACCGTGGCTAGAGCAGTTAGGGAAG 947
QY 820 CCACCTGCGCGGCGCCAGGAGGCTCAGCAGA 848
Db 948 CTACTGTGTTGCCAGTCACTGTGTTAA 976

RESULT 7
US-09-154-083-17
; Sequence 17, Application US/09154083
; Patent No. 6150513
; GENERAL INFORMATION:
; APPLICANT: Wu, Kai
; TITLE OF INVENTION: Polyketide Synthase Enzymes and Recombinant DNA
; FILE REFERENCE: 30062-20022.00 Polyketide Synthase Enz
; CURRENT APPLICATION NUMBER: US/09/154,083
; CURRENT FILING DATE: 1998-09-16
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
US-09-154-083-17

Query Match 3.2%; Score 42.8; DB 3; Length 466;
Best Local Similarity 46.4%; Pred. No. 0.2;
Matches 140; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 497 GGAGCGCTGTGGCGGTGTGAGAGGTGCCTGAAGCCTACGTGACATCCACACTGGCCT 556
Db 42 GGAGCGCGCGCGCGCGGTGAAGCGCTGCGGTGTCAGCCACGCTTCCACTCCGCGCT 101
QY 557 CAGTGGCAGTGGCGGTGCTGTGTGTCATTCTCCGAGGCGCTGCTGAAGAGCGGT 616
Db 102 CATGACGCGATGCTCGAGCGCTTCCGCCAGTTCGCCGAAGCGTGTCTGAGAGCGCCC 161
QY 617 CAAGATGGCATGCCACAGCGCTGGCCACCGCATCGTCCACAGACCCCTGCTGGGAC 676
Db 162 GCGGATCGGATCGTCTCGACCTGACCGCGCGTCTCGCGCGCGCGAGATCGGCAC 221
QY 677 GCCAAGATGCTGTGCACGAGGCGCAACACCCAGCCAGCTGCGCTCAGAGTGTGCAC 736
Db 222 GCCCGACTACTGGGTGGCGCAGTCCGCGAGTGTGCTAGCGTTCACAGCGGTCTGAGCG 281
QY 737 CCCGGTGGCACCACTATGAGTCCAGCCCTGGAGCAGGCGGCGCTGCGAGCAGC 796
Db 282 ACTGACGCGCAGGCGCTCGCACCTTCTGTGAGTGGCGCGCGCGAGTCTCTGTCGC 341
QY 797 CA 798
Db 342 CA 343

RESULT 8
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
```

```
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 3.2%; Score 42.2; DB 1; Length 7218;
Best Local Similarity 12.4%; Pred. No. 0.59;
Matches 17; Conservative 81; Mismatches 39; Indels 0; Gaps 0;

QY 837 GAGCTCAGCAAGTAGGTGGCTCTGGCCATCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 896
Db 1035 GAGCTTGGCTGAGGTGGAGGAGCTTGGATATATATATATATATATATATATATATAT 1094
QY 897 CTCCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTA 956
Db 1095 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1154
QY 957 CTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 973
Db 1155 YYYYYYYYYYYYYYYYYY 1171

RESULT 9
US-09-249-585A-2/c
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2
```


C:Accession: A41770
R:Dougherty, K.M.; Brandriss, M.C.; Valle, D.
J. Biol. Chem. 267, 871-875, 1992
A:Title: Cloning human pyrroline-5-carboxylate reductase cDNA by complementation in Saccharomyces cerevisiae
A:Reference number: A41770; MUID:921112821; PMID:1730675
A:Accession: A41770
A:Molecule type: mRNA
A:Residues: 1-319 <DOU>
A:Cross-references: GB:M77836; NID:g189497; PIDN:AAA36407.1; PID:g189498
A:Note: sequence extracted from NCBI backbone (NCBIN:75606, NCBIP:75608)
C:Genetics:
A:Gene: GDB:PYCRI; P5C: PYCR
A:Cross-references: GDB:135716; OMIM:179035
A:Map position: 17pter-17qter
C:Superfamily: pyrroline-5-carboxylate reductase
C:Keywords: oxidoreductase; proline biosynthesis

Query Match 42.2%; Score 584.5; DB 1; Length 319;
Best Local Similarity 46.4%; Pred. No. 2.2e-38;
Matches 124; Conservative 47; Mismatches 89; Indels 7; Gaps 3;

Qy 11 VGFVAGRMAGATAOGLIRACKVEAQHILASAPTDRLNLCFQAL--GCRTTHSNQEVLIQ 67
Db 3 VGFIFAGLAFALAKGFTAAAGYAAAHKIMASSP--DMDLATVSALRKMVGKLTPHNKETVQ 61
Qy 68 SCLLVIFATPHVLPVLAELVAPVVTTEHILSVAAAGVSLSTLEELP---PNTRVLRVL 124
Db 62 HSDVLFVAPVPHIIPFILDIGADIEDRHIVVSCAAGVTISSIEKKLSAFRPAVRICM 121
Qy 125 PNLPCVVQEGAIVMARGHRVGSSETKLLQHLLEACGRCEVPEAYVDIHTGLSGGVAFV 184
Db 122 TNPVVRREGATVYATGTHAQVEDGRLEQQLLSTVGFCTEVEEDLIDAVTGLSGSPAYA 181
Qy 185 CAPSEALAEAGVAKMGMPSSLAHRIAQTLLGTAKKMLLHEGHPAQLRSDVCTPGGTTIYG 244
Db 182 FTALDALADGGVKGMLPRRLRVLGAGALLGAAKMLLHSEQHPQLKDNVSSPGGATIIA 241
Qy 245 LHLEOGLLRAATMSAVEAATCRAKEL 271
Db 242 LHVLESGGFRSLINAVEASCIRTEL 268

RESULT 3
T06477
probable pyrroline-5-carboxylate reductase (EC 1.5.1.2) - garden pea
C:Species: Pisum sativum (garden pea)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T06477
R:Williamson, C.L.; Slocum, R.D.,
Plant Physiol. 100, 1464-1470, 1992
A:Title: Molecular cloning and evidence for osmoregulation of the dl-pyrroline-5-carboxylate reductase gene in Pisum sativum
A:Reference number: 215706
A:Accession: T06477
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-273 <WII>
A:Cross-references: EMBL:X62842; NID:g20850; PIDN:CAA44646.1; PID:g20851
A:Experimental source: cv. Wando
C:Genetics:
A:Gene: proc
C:Superfamily: pyrroline-5-carboxylate reductase
C:Keywords: oxidoreductase; proline biosynthesis

Query Match 36.3%; Score 502.5; DB 2; Length 273;
Best Local Similarity 39.7%; Pred. No. 4.7e-32;
Matches 104; Conservative 58; Mismatches 99; Indels 1; Gaps 1;

Qy 11 VGFVAGRMAGATAOGLIRACKVEAQHILASAPTDRLNLCFQALGCRTHSNQEVLOSCL 70
Db 13 LGFVAGKMAESIAKAGSRGVLFSRLVTAHSPRRAAFEISGITVLSNDDVVRASN 72
Qy 71 LVIFATPHVLPVLAELVAPVVTTEHILSVAAAGVSLSTLEELPPNTRVLRVLNPLPCV 130

RESULT 7
AE1124
1-pyrroline-5-carboxylate reductase (Proc) homolog lmo0396 [imported] - Lister
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AE1124
R:Glaser, P.; Francaul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.;
S.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournon,
O.K.; C. Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AE1077; MUID:21537279; PMID:11679669

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-266 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98475.1; PID:g16409774; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0396
C;Superfamily: pyrroline-5-carboxylate reductase

Qy	10	RVGFGVAGRMAGATAOGLIRAGKVEAQHILASAPTRNLCHEFOAL----	CGRTHSNQE	64
Dd	3	1R1F1G1AGMGTAMIRLQAQLANLVKREIIVGG--RNLEKLPLAEAFGLQITDTEK	59	
Qy	65	VLOSCLLVIFATKPHVLPVAVLAEVAPVVTTERHILSVAAAGVSLSTLEELLPPNTRVL	RVL	124
Dd	60	LEEQADIILIAVKPYTPIEILTYSKEKLTDPDKIIISVAAGVTIQDLEELTSAKTKIVKM	119	
Qy	125	NPLPCVOBGAIVMARGRHHVGSSTFKLLQHLLEACRCBEVPYAYVDIHTLUGSGVAFV	184	
Dd	120	PNTPALVGEAMSSVSPNTNYISEELKEVTAIFTSFGEAEVWSENLMDAVIGVSGSSPAY	179	
Qy	185	CAFSBALAEGAVKMGMPSSLAHRILAAQTLLGTAKMLLHEGOHPAQURSDVCTPGGTTIYG	244	
Dd	180	YMFIETALADGAVLGSMPRDRYAKFAQAVLGAAKTVLETGEHPGKLDMVTSFGGTTIEA	239	
Qy	245	LHALBOGGLRAATWSAVEEA	264	
Dd	240	VKSLENDFRSAVINAOCAA	259	

```

RESULT 8
AG1484
1-pyrroline-5-carboxylate reductase (proc) homolog lin0414 [imported] - Listeria
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AG1484
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.;
Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.;
Jones, L.M.; Karst, U.

```

Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1484

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-266 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC95647.1; PID:g16412843; GSPDB:GN00178

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: lin0414

C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 33.0%; Score 458; DB 2; Length 266;

Best Local Similarity 37.5%; Pred. No. 1.4e-28;

Matches 98; Conservative 58; Mismatches 97; Indels 8; Gaps 2;

QY 9 RRVGFGAGRWAGATAGGLIRACKVEAOHILASAPTRNLCHFQAL-----GCRTHSNQ 63

DB 2 KKIGFIGAGNNGAAMINGLAKANLKEEDILVCG---RDMEKLPKLTFFNGIQLTDTIA 58

QY 64 EVLQSCLLVIFATKPHVPAVLAEEVAPVVTTEHILVSAAGVSLSTLELLPPNTRVLRV 123

DB 59 QLAEQADIIILSVKPTTPEILTAVKDKITPEKIVISVAAGVTIKDEELTSTETKIVRV 118

QY 124 LPNLPVQVGAIYVARGHVGSSSEYKLLQHLLEACGRCEEVEPEAYVDIHTGLSGSVAF 183

DB 119 MPNTPALVGVAMSSISNPSNTAEETAHTISFSGKEAEVVAENLMDVAVGVGSSPAY 178

QY 184 VCAFSEALAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHGGHPAQLRSDVCTPGGTTTY 243

DB 179 VYMFIALADGAVLGNPRDKAYKFAQAQVLAAGAAKVVLETGHPGKLKMDVTPSPGGTTIE 238

QY 244 GLHLEOGLRAATMSAVEAA 264

DB 239 AVKSLEDTGFRSSVISAVQAA 259

RESULT 9

AG1867

pyrroline-5-carboxylate reductase [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

A:Accession: AG1867

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AG1867

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-270 <KUB>

A:Cross-references: GB:BA000019; PIDN:BA072446.1; PID:g17129833; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr0488

C:Superfamily: pyrroline-5-carboxylate reductase

Query Match

Best Local Similarity 32.9%; Score 455.5; DB 2; Length 270;

Matches 110; Conservative 43; Mismatches 109; Indels 7; Gaps 6;

QY 10 RVGFGAGRWAGATAGGLIRACKVEAOHILASAP--TDNRLCHFQALGCRTHSNQEVLOS 68

DB 4 KFLGLGGVGMGALLSLRIARGIYQPSVEIVSEPTARQAFLOQKYHGVGTTDNLVFTQ 63

QY 69 CLLLVTF-ATKPHVLPVLAEEVAPVVTTEH--ILVSAAGVSLSTLELLPPNTRVLRVLP 125

DB 64 AQDVVFLAVKPOVFAIAQELADTVTDHSPLVVSIAGVSLSQLEAAF--PQSPVIRAP 122

QY 126 NLPVQVGAIYVARGHVGSSSEYKLLQHLLEACGRCEEVEPEAYVDIHTGLSGSVAFVC 185

DB 123 NTPATVSGMTAICSGAYTAAHQKLAQIFSAVGEVVEVSESLMDAVTGLSGSGPAYVA 182

QY 186 AFSEALAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHGGQ-HPAQLRSDVCTPGGTTIYG 244

DB 183 LLVEALADGGVAGSLPRGIANQALQATVLGTAAQ-LLHESKLPALDKDRVTPSGGTTIAG 241

QY 245 LHLEOGLRAATMSAVEAAATCRAKELSR 273

DB 242 IAQLEKAGFRSALIEAVKATWRSQELGK 270

RESULT 10

REDEC

pyrroline-5-carboxylate reductase (EC 1.5.1.2) - *Escherichia coli* (strain K-12)C:Species: *Escherichia coli*

C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 01-Mar-2002

C:Accession: A00385; B64767

R:Deutch, A.H.; Smith, C.J.; Rushlow, K.E.; Kretscher, P.J.

Nucleic Acids Res. 10, 7701-7714, 1982

A:Title: *Escherichia coli* delta(1)-pyrroline-5-carboxylate reductase: gene sequence,

A:Reference number: A00385; MUID:83116986; PMID:6296787

A:Accession: A00385

A:Molecule type: DNA

A:Residues: 1-269 <DEU>

A:Cross-references: GB:J01665; NID:g147358; PIDN:AAA86433.1; PID:g147359

A:Note: parts of this sequence, including the amino and carboxyl ends of the mature p

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B64767

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-269 <BLAT>

A:Cross-references: GB:AE000145; GB:U00096; NID:g1786580; PIDN:AAC73489.1; PID:g17865

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: proC

A:Map position: 9 min

C:Function:

A:Description: catalyzes reduction of pyrroline-5-carboxylate to proline

A:Pathway: proline biosynthesis

A:Note: third enzyme in the proline biosynthetic pathway

C:Superfamily: pyrroline-5-carboxylate reductase

C:Keywords: oxidoreductase; proline biosynthesis

Query Match 30.6%; Score 423.5; DB 1; Length 269;

Best Local Similarity 34.6%; Pred. No. 7e-26;

Matches 92; Conservative 56; Mismatches 117; Indels 1; Gaps 1;

QY 9 RRVGFGAGRWAGATAGGLIRACKVEAOHILASAPT--DRNLCHFQALGCRTHSNQEVLO 67

DB 3 KKIGFIGCGNMKAILGLLIASQVLPQIWTYTPSPDKVAALHDPQFGINAESAQEAQ 62

QY 68 SCLLVIFATKPHVLPVLAEEVAPVVTTEHILVSAAGVSLSTLELLPPNTRVLRVLPNL 127

DB 63 IADITFAAVKPGIMIKVLSEITSSLNKOSLVSIAGVTLQDLARALGHDKRIIRAMPNT 122

QY 128 PCVQVGAIYVARGHVGSSSEYKLLQHLLEACGRCEEVEPEAYVDIHTGLSGSVAFVCA 187

DB 123 PALVNMAGMTSVTPNALVTPEDTADVLNIFRCFGEAEVIAEPMIHPVVGSGSPAYVFM 182

QY 188 SEALAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHGGHPAQLRSDVCTPGGTTIYGLHA 247

DB 183 IEAMADAAVLGMPRAQAYKFAAQAVMGSAKVVLETGEHPGALKDKMVKSPGGTTIEAVRV 242

QY 248 LEQGLRAATMSAVEAAATCRAKELSR 273

DB 243 LEKGFRAAVIEAMTKMEKSEKLSK 268

RESULT 11

D90683
pyrroline-5-carboxylate reductase [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: D90683
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90683
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033859.1; PID:g13359893; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs0437
C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 30.6%; Score 423.5; DB 2; Length 269;
Best Local Similarity 34.6%; Pred. No. 7e-26;
Matches 92; Conservative 56; Mismatches 117; Indels 1; Gaps 1;
QY 9 RRVGFGAGRMAGATAGLIRAGKVEAQHILASAPT-DRNLCHFQALGCRTHSHNOEVLQ 67
Db 3 KKGIFGCGNMGKAILGLIASQVLPQGIWVTPSPDKVAALHDKFGINAEASQVAF 62
QY 68 SCLLVIFATKPHVLPVLAEAVPVVTTTEHILSVAAAGVSLSTLEELLPPNTRVLRVLPNL 127
Db 63 IADIIFAAVKPGIMIKVLSEITSSLNKDSLVSIAAGVTLDQALRALGHDRIIRAMPNT 122
QY 128 PCVQVEGAIVMARGHVSSEFKLQHLLEACGCEVEPEAYVDIHTGLSGGVAFVCAF 187
Db 123 PALVNAGMTSVTPNALVTPEDADVLNIFRCFGEAEVIAEPMIHPVWVGSGSPAYVFMF 182
QY 188 SEALAEGAVKMGMPSSLAHRIAATAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGLHA 247
Db 183 IEAMADAAYLGMCPRAQAYKFAAQAVMGSAKMWLETEGHPGALKDMVCSFGGTTIEAVRV 242
QY 248 LEQGLRAATMSAVEAATCRAKELSR 273
Db 243 LEEKGFRAAVIEAMTKCKMEKSEKLSK 268

RESULT 12

H85533
pyrroline-5-carboxylate reductase [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85533
R:Ferna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85533
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <STO>
A:Cross-references: GB:AE005174; NID:g12513226; PIDN:AAG54732.1; GSPDB:GN00145; UWG:Z04
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: PROC
C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 30.6%; Score 423.5; DB 2; Length 269;
Best Local Similarity 34.6%; Pred. No. 7e-26;
Matches 92; Conservative 56; Mismatches 117; Indels 1; Gaps 1;

QY 9 RRVGFGAGRMAGATAGLIRAGKVEAQHILASAPT-DRNLCHFQALGCRTHSHNOEVLQ 67
Db 3 KKGIFGCGNMGKAILGLIASQVLPQGIWVTPSPDKVAALHDKFGINAEASQVAF 62
QY 68 SCLLVIFATKPHVLPVLAEAVPVVTTTEHILSVAAAGVSLSTLEELLPPNTRVLRVLPNL 127
Db 63 IADIIFAAVKPGIMIKVLSEITSSLNKDSLVSIAAGVTLDQALRALGHDRIIRAMPNT 122
QY 128 PCVQVEGAIVMARGHVSSEFKLQHLLEACGCEVEPEAYVDIHTGLSGGVAFVCAF 187
Db 123 PALVNAGMTSVTPNALVTPEDADVLNIFRCFGEAEVIAEPMIHPVWVGSGSPAYVFMF 182
QY 188 SEALAEGAVKMGMPSSLAHRIAATAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGLHA 247
Db 183 IEAMADAAYLGMCPRAQAYKFAAQAVMGSAKMWLETEGHPGALKDMVCSFGGTTIEAVRV 242
QY 248 LEQGLRAATMSAVEAATCRAKELSR 273
Db 243 LEEKGFRAAVIEAMTKCKMEKSEKLSK 268

RESULT 13

T36286
pyrroline-5-carboxylate reductase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T36286
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, May 1998
A:Reference number: 221603
A:Accession: T36286
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-284 <SEE>
A:Cross-references: EMBL:AL049819; PIDN:CAB42663.1; GSPDB:GN00070; SCOEDB:SCE7.04c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: proc; SCOEDB:SCE7.04c
C:Superfamily: pyrroline-5-carboxylate reductase

Query Match 30.3%; Score 420.5; DB 2; Length 284;
Best Local Similarity 35.4%; Pred. No. 1.3e-25;
Matches 95; Conservative 53; Mismatches 119; Indels 1; Gaps 1;

QY 5 EPSPRRVGFGAGRMAGATAGLIRAGKVEAQHILASAPTDRNLCHFQALGCRTHSHNOE 64
Db 13 EPMTQKVAVLTGKIGKIGALLSGMIGAGWAPAD-LVTARRRERADELRHGVTPVTNAE 71
QY 65 VLQSLLVIFATKPHVLPVLAEAVPVVTTTEHILSVAAAGVSLSTLEELLPPNTRVLRVL 124
Db 72 AAKAADTLTLTVKPDQMDTLDELAPHVPADRLVISGAAGVPTSFEEERLAPGTPVVRVM 131
QY 125 PNLPCVQVEGAIVMARGHVSSEFKLQHLLEACGCEVEPEAYVDIHTGLSGGVAFV 184
Db 132 TMTALVDEAMSVISAGTATAAHLTHTTEIFGAVGKTLRVPESEQOQACTALSQSPAYF 191
QY 185 CAFSEALAEAGVAKMGPSSLAHRIAATAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIY 244
Db 192 FYLYEAMTADAGILGLPRDKAHDILVQSAIGAAKMLRDSGEHPVKLRNVTSPAGTTINA 251
QY 245 LHALEQGLRAATMSAVEAATCRAKELS 272
Db 252 IRELENHGVRAALIAALEAARDRSRELA 279

RESULT 14

AI0549
pyrroline-5-carboxylate reductase [imported] - Salmonella enterica subsp. enterica se
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AI0549
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:53:21 ; Search time 22 Seconds
(without alignments)
1479.099 Million cell updates/sec

Title: US-09-806-536A-14

Perfect score: 1386

Sequence: 1 MAAPSPRRVGVGAGRMA.....AATNSAVEATCRAKELSRK 274

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA.*

- 1: /cgn2.6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 2: /cgn2.6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep.*
- 3: /cgn2.6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2.6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2.6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep.*
- 6: /cgn2.6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2.6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2.6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.1*
- 11: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.2*
- 12: /cgn2.6/ptodata/1/pubpaa/US09_NEW_PUB.pep.3*
- 13: /cgn2.6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 14: /cgn2.6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 15: /cgn2.6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2.6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 17: /cgn2.6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1386	100.0	274	15	US-10-161-418A-11
2	1386	100.0	274	15	US-10-161-418A-13
3	584.5	42.2	319	15	US-10-161-418A-10
4	584.5	42.2	319	15	US-10-161-418A-12
5	560.5	40.4	315	10	US-09-912-717-3
6	537.5	38.8	314	10	US-09-912-717-1
7	477	34.4	319	15	US-10-161-418A-14
8	467.5	33.7	255	11	US-09-925-300-1218
9	408.5	29.5	269	15	US-10-156-761-12258
10	401	28.9	284	15	US-10-128-714-3252
11	401	28.9	284	15	US-10-128-714-3252
12	366	26.2	295	11	US-09-712-363-174
13	362.5	26.2	270	11	US-09-738-626-3960
14	116	8.4	144	10	US-09-939-980-445
15	114	8.2	299	10	US-09-912-020-256
16	95.5	6.9	549	15	US-10-156-761-14029

17	95	6.9	807	12	US-09-930-020A-2	Sequence 2, Appli
18	93.5	6.7	289	15	US-10-156-761-9583	Sequence 9583, Ap
19	93	6.7	2436	10	US-09-795-693-8	Sequence 8, Appli
20	93	6.7	2436	15	US-10-156-239-8	Sequence 8, Appli
21	93	6.7	2436	15	US-10-199-485-8	Sequence 8, Appli
22	92	6.6	408	11	US-09-712-363-273	Sequence 273, App
23	91.5	6.6	452	15	US-10-156-761-12493	Sequence 12493, A
24	89.5	6.5	323	15	US-10-156-761-9372	Sequence 9372, Ap
25	89.5	6.5	336	15	US-10-156-761-10217	Sequence 10217, A
26	89.5	6.5	488	15	US-10-156-761-7949	Sequence 7949, Ap
27	89.5	6.5	858	10	US-09-815-242-11396	Sequence 11396, A
28	89	6.4	447	15	US-10-156-761-9880	Sequence 9880, Ap
29	87.5	6.3	1024	15	US-10-211-962-44	Sequence 44, Appli
30	87.5	6.3	6145	15	US-10-156-761-7962	Sequence 7962, Ap
31	87	6.3	426	15	US-10-156-761-8581	Sequence 8581, Ap
32	87	6.3	949	10	US-09-841-835-10	Sequence 10, Appli
33	87	6.3	1327	10	US-09-841-835-2	Sequence 2, Appli
34	87	6.3	1327	12	US-09-972-115A-8	Sequence 8, Appli
35	86.5	6.2	340	10	US-09-815-242-10852	Sequence 10852, A
36	86.5	6.2	342	10	US-09-815-242-4924	Sequence 4924, Ap
37	86.5	6.2	1835	15	US-10-156-761-7963	Sequence 7963, Ap
38	86.5	6.2	4809	15	US-10-156-761-9090	Sequence 9090, Ap
39	86	6.2	492	10	US-09-815-242-11079	Sequence 11079, A
40	86	6.2	499	15	US-10-156-761-9271	Sequence 9271, Ap
41	85.5	6.2	342	15	US-10-156-761-14291	Sequence 14291, A
42	85.5	6.2	748	10	US-09-815-242-12792	Sequence 12792, A
43	85.5	6.2	792	10	US-09-815-242-12327	Sequence 12327, A
44	85.5	6.2	1687	15	US-10-094-679-3	Sequence 3, Appli
45	85	6.1	1074	10	US-09-509-196A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-161-418A-11
; Sequence 11, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-11

Query Match	100.0%	Score 1386;	DB 15;	Length 274;
Best Local Similarity	100.0%	Pred. No. 1.5e-132;		
Matches 274;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAAAPSPRRVGVGAGRMAIAOGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTH	60	
Db	1	MAAAPSPRRVGVGAGRMAIAOGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTH	60	
QY	61	SNOEVLQSCLLVIFATKPHVLPVLAIEVAPVVTTTHILSVAAAGVSLSTLELLPPNTRV	120	
Db	61	SNOEVLQSCLLVIFATKPHVLPVLAIEVAPVVTTTHILSVAAAGVSLSTLELLPPNTRV	120	
QY	121	LRVLNPLPCVVOEGAIVMARGHRVGSSETKLQHLLEACGRCEEVEPYVDIHTGLSGG	180	
Db	121	LRVLNPLPCVVOEGAIVMARGHRVGSSETKLQHLLEACGRCEEVEPYVDIHTGLSGG	180	
QY	181	VAFVCAFEALAEAGAVKMGMPSSLAHRTAAQTLLGTAKMLLHEGHPAQLRSDVCTPGT	240	

```

Db 181 VAFVCAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALEQGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALEQGLRAATMSAVEAATCRAKELSRK 274

RESULT 2
US-10-161-418A-13
; Sequence 13, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; PRIOR FILING DATE: 2002-09-11
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-13

Query Match 100.0%; Score 1386; DB 15; Length 274;
Best Local Similarity 100.0%; Pred. No. 1.5e-132;
Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAEPPSPRRVGFVAGRMAGTAQGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTH 60
Db 1 MAAEPPSPRRVGFVAGRMAGTAQGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTH 60

QY 61 SNOEVLOSCLLVIFATKPHVLPVLAIEVAPVVTTEHILVSVAAAGVSLSTLEELPPNTRV 120
Db 61 SNOEVLOSCLLVIFATKPHVLPVLAIEVAPVVTTEHILVSVAAAGVSLSTLEELPPNTRV 120

QY 121 LRVLPNLPCVQVEGAIVMARGHRVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGS 180
Db 121 LRVLPNLPCVQVEGAIVMARGHRVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGS 180

QY 181 VAFVCAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240

QY 241 TIYGLHALEQGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALEQGLRAATMSAVEAATCRAKELSRK 274

RESULT 3
US-10-161-418A-10
; Sequence 10, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; PRIOR FILING DATE: 2002-09-11
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-10

Query Match 42.2%; Score 584.5; DB 15; Length 319;
Best Local Similarity 46.4%; Pred. No. 4.5e-51;
Matches 124; Conservative 47; Mismatches 89; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGTAQGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTHSNOEVLO 67
Db 3 VGFVAGRMAGTAQGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTHSNOEVLO 67

QY 68 SCLLVIFATKPHVLPVLAIEVAPVVTTEHILVSVAAAGVSLSTLEELP---PNTYRLVRL 124
Db 62 HSDVFLAVKPHIIPFILDIGADIEDRHIVVSCAAGVTISSIEKKLSAFRPAPIVRCM 121

QY 125 PNLPCVQVEGAIVMARGHRVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGSVAFV 184
Db 122 TNPFPVVVREGATVYATGTHAQVEDGRMBEQLLSTVGFCTEVEDLIDAVTGLSGSGPAYA 181

QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGQHPAQLRSDVCTPGGTIYG 244
Db 182 FTALDALADGGVKGMLPRRLAVRLGAQALLGAAKMLLHSEQHPQLKDNVSSPGGATTHA 241

QY 245 LHALEQGLRAATMSAVEAATCRAKEL 271
Db 242 LHVLESGGFRSLLINAVEASCINTREL 268
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-10

Query Match 42.2%; Score 584.5; DB 15; Length 319;
Best Local Similarity 46.4%; Pred. No. 4.5e-51;
Matches 124; Conservative 47; Mismatches 89; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGTAQGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTHSNOEVLO 67
Db 3 VGFVAGRMAGTAQGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTHSNOEVLO 67

QY 68 SCLLVIFATKPHVLPVLAIEVAPVVTTEHILVSVAAAGVSLSTLEELP---PNTYRLVRL 124
Db 62 HSDVFLAVKPHIIPFILDIGADIEDRHIVVSCAAGVTISSIEKKLSAFRPAPIVRCM 121

QY 125 PNLPCVQVEGAIVMARGHRVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGSVAFV 184
Db 122 TNPFPVVVREGATVYATGTHAQVEDGRMBEQLLSTVGFCTEVEDLIDAVTGLSGSGPAYA 181

QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGQHPAQLRSDVCTPGGTIYG 244
Db 182 FTALDALADGGVKGMLPRRLAVRLGAQALLGAAKMLLHSEQHPQLKDNVSSPGGATTHA 241

QY 245 LHALEQGLRAATMSAVEAATCRAKEL 271
Db 242 LHVLESGGFRSLLINAVEASCINTREL 268

RESULT 4
US-10-161-418A-12
; Sequence 12, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; PRIOR FILING DATE: 2002-09-11
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-161-418A-12

Query Match 42.2%; Score 584.5; DB 15; Length 319;
Best Local Similarity 46.4%; Pred. No. 4.5e-51;
Matches 124; Conservative 47; Mismatches 89; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGTAQGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTHSNOEVLO 67
Db 3 VGFVAGRMAGTAQGLIRAGKVEAQHILASAPTDNRNLCHFQALGCRTHSNOEVLO 67

QY 68 SCLLVIFATKPHVLPVLAIEVAPVVTTEHILVSVAAAGVSLSTLEELP---PNTYRLVRL 124
Db 62 HSDVFLAVKPHIIPFILDIGADIEDRHIVVSCAAGVTISSIEKKLSAFRPAPIVRCM 121

QY 125 PNLPCVQVEGAIVMARGHRVGSSETKLLQHLLEACGRCEEPVPEAYVDIHTGLSGSVAFV 184
Db 122 TNPFPVVVREGATVYATGTHAQVEDGRMBEQLLSTVGFCTEVEDLIDAVTGLSGSGPAYA 181

QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGQHPAQLRSDVCTPGGTIYG 244
Db 182 FTALDALADGGVKGMLPRRLAVRLGAQALLGAAKMLLHSEQHPQLKDNVSSPGGATTHA 241

QY 245 LHALEQGLRAATMSAVEAATCRAKEL 271
Db 242 LHVLESGGFRSLLINAVEASCINTREL 268
```

Db 242 LHVLESGGFRSLINAVEASCIRTREL 268

RESULT 5

US-09-912-717-3

; Sequence 3, Application US/09912717

; Patent No. US20020081691A1

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; Corley, Neil C.

; Baughn, Mariah R.

; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE

; HOMOLOG

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/912,717

; FILING DATE: 24-Jul-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/565,910

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Cerrone, Michael C

; REGISTRATION NUMBER: 39,132

; REFERENCE/DOCKET NUMBER: PF-0532 US

; TELEPHONE: 650-855-0555

; TELEFAX: 650-855-0572

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 315 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GENBANK

; CLONE: 189498

; SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-912-717-3

Query Match 40.4%; Score 560.5; DB 10; Length 315;

Best Local Similarity 46.1%; Pred. No. 1.2e-48;

Matches 123; Conservative 46; Mismatches 87; Indels 11; Gaps 5;

QY 11 VGFVAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNCHFOAL---GCRTHSNQEVQLQ 67

Db 3 VGFAGQL--AFAGFTAAGVLAHAKIMASSP-DMDLATYSALKRMGVKLTLPNKETVQ 59

QY 68 SCLLVIFATKPHVLPVLAIEAPVVTTEHILVSVAGVSLSTLELLP---PNTRVLRVL 124

Db 60 HSDVLFVAVKPHIIPFLDEIGADIEDRHVIVSCAAGVTISSIEKKLSAFRPAPIVRCM 119

QY 125 PNLPCVQVQGAIVMARGHVGSSSETKLLQHLLEACGCEEVPEAYVDIHTGLSGGVAFV 184

Db 120 TNPVVVREGAVYATGTHAQVEDQLMEQLLSTVGFCTEVEDLIDAVTGLSGGPAY- 178

QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGTTIYG 244

Db 179 -AFTALDADGGVKGMLPRELAVRLCAQLLGNAAKMLLHSEHQPGQLKNVSPGGATIHA 237

QY 245 LHALEGGGLRAATMSAVEAATCRAKEL 271

Db 238 LHVLESGGFRSLINAVEASCIRTREL 264

RESULT 6

US-09-912-717-1

; Sequence 1, Application US/09912717

; Patent No. US20020081691A1

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; Corley, Neil C.

; Baughn, Mariah R.

; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE

; HOMOLOG

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/912,717

; FILING DATE: 24-Jul-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/565,910

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Cerrone, Michael C

; REGISTRATION NUMBER: 39,132

; REFERENCE/DOCKET NUMBER: PF-0532 US

; TELEPHONE: 650-855-0555

; TELEFAX: 650-855-0572

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 314 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: PROSNON01

; CLONE: 2278458

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-912-717-1

Query Match 38.8%; Score 537.5; DB 10; Length 314;

Best Local Similarity 45.3%; Pred. No. 2.6e-46;

Matches 121; Conservative 45; Mismatches 88; Indels 13; Gaps 5;

QY 11 VGFVAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNCHFOAL---GCRTHSNQEVQLQ 67

Db 3 VGFAGQL--LAFRTAAGILSAHKIIASSP-EMNLPTVLSALKRMGVNLTNSKKTIVK 57

QY 68 SCLLVIFATKPHVLPVLAIEAPVVTTEHILVSVAGVSLSTLELLP---PNTRVLRVL 124

Db 58 HSDVLFVAVKPHIIPFLDEIGADVQARHIVSCAAGVTISSVEKKLMFQAPKIVRCM 117

QY 125 PNLPCVQVQGAIVMARGHVGSSSETKLLQHLLEACGCEEVPEAYVDIHTGLSGGVAFV 184

Db 118 TNPVVVQEGAVYATGTHALVEDQGLLEQLMSSVGCTEVEDLIDAVTGLSGGPAY- 176

QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGTTIYG 244

```

US-09-925-300-1218
; Sequence 1218, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1218
; LENGTH: 255

```

Db 174 PAYFFYLVAMTDAGILLGPRDKAHLIVQSAIGAAIMLRDSEGHVPKLRNWTSPAGT 233
QY 241 TIYGLHALEOGLRAATMSAVEAATCRAKELS 272
Db 234 TINARELENHGVRAALIALEAARDRSRELA 265

RESULT 10

US-10-128-714-3252
; Sequence 3252, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3252
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3252

Query Match 28.9%; Score 401; DB 15; Length 284;
Best Local Similarity 37.3%; Pred. No. 1.6e-32;
Matches 101; Conservative 45; Mismatches 115; Indels 10; Gaps 4;

QY 10 RVGFVAGRMAGAIAGQLIRAGKVEAQHILASAPTRNLCHFOALGCRTHSNQEVQSC 69
Db 8 KLAFIGGNNMASAIIGLV-SQDMNPANITVSEPDVNRKIAKLGVTQTTSSNGEAAAANA 66
QY 70 LLVIFATKPHVLPVLAELAVPV---VTTEHILVSAAGVSLSTLELLPPN---TRVLR 122
Db 67 DIVIIAVKPTTKNVCELATANSORTSLPVVVSIAGITLNSMKEWLRTNDGRTAHIVR 126
QY 123 VLPNLCVVOEGAIVMARGHVGSSSETKLLQHLLEACGRCEE--VPEAYVDIHTGLSGSG 180
Db 127 VMPNTPALVKEGASGLASDDVTAEKELIGALLOSVS KATEWVEKEELDDVVTGLSGSG 186
QY 181 VAFVCAFEALAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 187 PAYFFAMVEHLVASATALGSLKEQATRLAAQTCLGAGKMLVESSEPAQLRKNVTSNPGT 246
QY 241 TIYGLHALEOGLRAATMSAVEAATCRAKEL 271
Db 247 THAALQTTFESLNFEKIVDKAVQAATSRRAEL 277

RESULT 11

US-10-128-714-8252
; Sequence 8252, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo

; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8252
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8252

Query Match 28.9%; Score 401; DB 15; Length 284;
Best Local Similarity 37.3%; Pred. No. 1.6e-32;
Matches 101; Conservative 45; Mismatches 115; Indels 10; Gaps 4;

QY 10 RVGFVAGRMAGAIAGQLIRAGKVEAQHILASAPTRNLCHFOALGCRTHSNQEVQSC 69
Db 8 KLAFIGGNNMASAIIGLV-SQDMNPANITVSEPDVNRKIAKLGVTQTTSSNGEAAAANA 66
QY 70 LLVIFATKPHVLPVLAELAVPV---VTTEHILVSAAGVSLSTLELLPPN---TRVLR 122
Db 67 DIVIIAVKPTTKNVCELATANSORTSLPVVVSIAGITLNSMKEWLRTNDGRTAHIVR 126
QY 123 VLPNLCVVOEGAIVMARGHVGSSSETKLLQHLLEACGRCEE--VPEAYVDIHTGLSGSG 180
Db 127 VMPNTPALVKEGASGLASDDVTAEKELIGALLOSVS KATEWVEKEELDDVVTGLSGSG 186
QY 181 VAFVCAFEALAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 187 PAYFFAMVEHLVASATALGSLKEQATRLAAQTCLGAGKMLVESSEPAQLRKNVTSNPGT 246
QY 241 TIYGLHALEOGLRAATMSAVEAATCRAKEL 271
Db 247 THAALQTTFESLNFEKIVDKAVQAATSRRAEL 277

RESULT 12

US-09-712-363-174
; Sequence 174, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,

```
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-174

Query Match          26.4%; Score 366; DB 11; Length 295;
Best Local Similarity 31.6%; Pred. No. 6.1e-29;
Matches 90; Conservative 55; Mismatches 116; Indels 24; Gaps 3;

QY 10 RVGFVAGRMAGAGIAGLLRAGKVEAQHILASAPTRDNLCHFOALGCRTHSNQEVLOS 69
DB 7 RTAIIIGSGSIGALLSGLLRAGQVKDLVVAERMPDRANYLAQTSYSLVT-SAADAVENA 65
QY 70 LLVIFATKPHVLPVLAIEVAPV-----TTEHILSVSVAAGVSLSTLELLPPNTRVLRV 123
DB 66 TFVVAVKPADVEPIVLAIDLANATAAENDSAEQVFVVVAGITIAEFESKLPAGTPVRA 125
QY 124 LPNLPVQVEGAIIVMARGHVGSSETKLLQHLLEACGRCVEEVPAYVDIHTGLSGGVAF 183
DB 126 MPNAAALVGAVTALAKRFVTPQOLEVSALFDVAGVGLTVPESQLDAVAVSGSPAY 185
QY 184 VCAFSEALAEAGVAKMGMPSSLAHRTAAQTLLGTAKMLLHEGQH----- 226
DB 186 FFLVLEALVDAGVGLSRQVATDLAAQTMAAGSAMLLERMEDOGGANGELMGLRVDLT 245
QY 227 PAQLRSDVCTPGGTTTLYGLHALEOGLRAATMSAVEAATCRAKEL 271
DB 246 ASRLRAAVTSPGGTTAAALRELERGGFRMAVDAVAQAQAKSRSEQL 290

RESULT 13
US-09-738-626-3960
; Sequence 3960, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARGHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3960
; LENGTH: 270

; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3960

Query Match          26.2%; Score 362.5; DB 11; Length 270;
Best Local Similarity 30.6%; Pred. No. 1.2e-28;
Matches 83; Conservative 63; Mismatches 114; Indels 11; Gaps 4;

QY 11 VGFVAGRMAGAGIAGLLRAGKVEAQHILASAPTRDNLCHFOALGCRTHSNQEV 66
DB 4 IAVIGGGQIGFALVSLI-AANMNPQIRV---TNRSEERGQELRDYRGLNMTDNSQAA 59
QY 67 QSCLLVIFATKPHVLPVLAIEVAPV---TTEHILSVSVAAGVSLSTLELLPPNTRVLRV 123
DB 60 DEADVFLCVKPKFIVEVLSEITGTLDNNSAQSVVSMAGISIAAAMEESASAGLPVVRV 119
QY 124 LPNLPVQVEGAIIVMARGHVGSSETKLLQHLLEACGRCVEEVPAYVDIHTGLSGGVAF 183
DB 120 MPNTPMLVGKGMSTVTKGRYVDAEQLEQVKDLSTVGDVLEVAESDIDATAMSGSPAY 179
QY 184 VCAFSEALAEAGVAKMGMPSSLAHRTAAQTLLGTAKMLLHEGQHQAQLRSDVCTPGGTTT 243
DB 180 LFLVTEALIEAGVNLGLPRATAKKLAVASPEGAATMMKETGKPPSELRAAGVSSPAGTTVA 239
QY 244 GLHLEOGLRAATMSAVEAATCRAKELSRK 274
DB 240 AIRELEESGIRGAFYRAAQACADRSEELGKR 270

RESULT 14
US-09-939-980-445
; Sequence 445, Application US/09939980
; Patent No. US20020082234A1
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; Burnham, Martin
; Hodgson, John
; Knowles, David
; Lonetto, Michael
; Nicholas, Richard
; Pratt, Julie
; Reichard, Richard
; Rosenberg, Martin
; Ward, Judith
; TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,
; Polypeptides and Their Uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/939,980
; FILING DATE: 27-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/936,165
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
```


Mon Jul 28 08:07:03 2003

```

QY      214 L-CTAKMLLHEGQHPA-----QLRSDVCTPGG-----TTIYGLHALFQGLRA 255
Db      220 LDRAKAPMVMDFRNFKPGFRIDJHIKDLANALDTSHGVGAQLPLTAAVMMQALRADGLGT 279
QY      256 ATMSAVEAATCRAKELSR 273
Db      280 ADHSAL---ACYYEKLAK 294

```

Search completed: July 21, 2003, 09:57:02
Job time : 24 secs

Search completed: July 21, 2003, 09:57:02
Job time : 24 secs

```

;       TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 445:
;
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 144 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;
;   MOLECULE TYPE: Protein
;
;   SEQUENCE DESCRIPTION: SEQ ID NO: 445:
;
US-09-939-980-445

      8.4%;   Score 116;   DB 10;   Length 144;
Best Local Similarity 26.8%;   Pred NO. 0.00054;
Matches 30;   Conservative 24;   Mismatches 56;   Indels 2;   Gaps 2;

16  AGRMAGIAAGLLTRACKVEAQHI -TASAPTDRLNLCHF-QALGCRTHHSNQEVLSCLLVI 73
   || || || || || || || || || || || || || || || || || || || || ||
12  AGNMAQAIFGTGINSNDLANDIYLTNKSNEQALAKAFKLGVNYSYDDATLLKADADYVF 71

74  FATKPHPLPAVLAEVAPVWTTTHILYSVAAGVSLSTLELLPPNTRVLRLVLP 125
   || || || || || || || || || || || || || || || || || || || || ||
72  IGTKPHDFEDALATRIKPHITTKDKCFMSIAGPTDVIYXOOLECONXPARTWP 123

```

```

RESULT 15
US-09-912-020-256
; Sequence 256, Application US/09912020
; Patent No. US2002004592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001DVI
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 256
; LENGTH: 299
; TYPE: PRT
; ORGANISM: E. Coli
US-09-912-020-256

```

	Query Match	8.2%	Score 114;	DB 10;	Length 299;
	Best Local Similarity	21.7%;	Pred. NO. 0.0024;		
	Matches 69;	Conservative 46;	Mismatches 119;	Indels 84;	Gaps 16;
QY	10	RVGVFGARMAGATAOGLIRAGKVEAQHILASAPTRDN---	LCHPQALGCRTHUSNOEVL	66	
DB	7	KVGFILGILGKMPKNNLLKAG-----	YSLVADRNPETAIVIAAGATASTAKAIA	59	
QY	67	QSC--LLVIFATKPHVLPAVLAEPVVTTEHILVSVAAGSVLSLSTLELLPNPTRVLRVL	124		
DB	60	BQCDVIITMLPNSPHVKEVALGE-----	NGIIIEAKPGTVLIDMSSIAPLASREI---	109	
QY	125	PNLPCVQOEGAIVNARG-----	RHVGSSETKILQHLLEA-CGRCEVPEAYVD-----	171	
DB	110	-----	SEALKAGIDMLDAPVGGSGPKAIDGTVLSVWVGDKAIAFDKYIDYLMKAMAG	160	
QY	172	-----	IHTGLSGSG-----	VAF-VCAPEALAECAVKMGMPSSLAHR-----	IAAQTL 213
DB	161	SVWHTGEIGAGNVTKLANOVIVAINLTAAMSEALT-	ATKAGNPDLVTVQAIRGGLAGSTV	219	

THE FRODO BAGGINS

FT Modified-site 259 /note= "potential phosphorylation site"
 FT Modified-site 265 /note= "potential phosphorylation site"
 FT XX
 XX
 PN WO200020604-A2.
 XX 13-APR-2000.
 XX
 PF 06-OCT-1999; 99WO-US23434.
 XX
 PR 06-OCT-1998; 98US-0172227.
 PR 02-DEC-1998; 98US-0155202.
 PR 10-MAR-1999; 99US-0123911.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 PI Lal P, Guegler KJ, Gorgone GA, Corley NC, Baughn MR, Tang YT;
 PI Hillman JL, Bandman O, Azimzai Y, Au-Young J, Yue H, Lu DAM;
 PI Yang J;
 XX
 DR WPI; 2000-303785/26.
 DR N-PSDB; AAA09388.
 XX
 PT Purified polypeptide for treating or preventing disorders associated
 PT with decreased expression or activity of oxidoreductase molecules
 XX
 PS Claim 1; Page 84; 97pp; English.
 XX
 CC This OXRE-14 has identity with pyrroline-5-carboxylate-reductase.
 CC The polypeptides are useful for treating or preventing a disorder
 CC associated with decreased expression or activity of OXRE. Antagonists of
 CC OXRE are useful for treating or preventing a disorder associated with
 CC increased expression or activity of OXRE. The disorders include cell
 CC proliferative disorders (cirrhosis, hepatitis), cancer (leukemia,
 CC melanoma), hypopituitarism and hyperpituitarism, hypothyroidism and
 CC hyperthyroidism, metabolic disorders (Addison's disease, cystic
 CC fibrosis), reproductive disorders (infertility, ovulatory defects),
 CC neurological disorders (Alzheimer's disease, Parkinson's disease,
 CC multiple sclerosis), mental disorders (anxiety, schizophrenia),
 CC autoimmune/inflammatory disorders (acquired immunodeficiency syndrome
 CC (AIDS), asthma, osteoarthritis), and viral infections. The
 CC polynucleotides may be used in Southern or Northern analysis, polymerase
 CC chain reaction (PCR), or in enzyme-linked immunosorbent assays (ELISA).
 XX
 SQ Sequence 274 AA;
 Query Match 100.0%; Score 1386; DB 21; Length 274;
 Best Local Similarity 100.0%; Pred. No. 5.7e-131;
 Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAAEPSPRVGFVAGRMAGATAOGLIRAGKVEAQHILASAPTRNLCHFQALGCRTH 60
 DB 1 MAAAEPSPRVGFVAGRMAGATAOGLIRAGKVEAQHILASAPTRNLCHFQALGCRTH 60
 QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEAPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120
 DB 61 SNOEVLQSCLLVIFATKPHVLPVLAEAPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120
 QY 121 LRVLPNLPVQVQEGALVMAGRHVSSETKLLQHLLEACGRCVEEPEAYVDIHTGLSGG 180
 DB 121 LRVLPNLPVQVQEGALVMAGRHVSSETKLLQHLLEACGRCVEEPEAYVDIHTGLSGG 180
 QY 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAATQLLGTAKMLLHEGHPAQLRSDVCTPGT 240
 DB 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAATQLLGTAKMLLHEGHPAQLRSDVCTPGT 240
 QY 241 TIYGLHALQGGGLRAATNSAVEAATCRAKELSRK 274
 DB 241 TIYGLHALQGGGLRAATNSAVEAATCRAKELSRK 274

RESULT 2

AAB95591
 ID AAB95591 standard; Protein; 274 AA.
 AC AAB95591;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ.ID NO:18269.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN BP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 XX full-length cDNAs.
 PS Claim 8; SEQ ID 18269; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 SQ Sequence 274 AA;
 Query Match 100.0%; Score 1386; DB 22; Length 274;
 Best Local Similarity 100.0%; Pred. No. 5.7e-131;
 Matches 274; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAAAEPSPRVGFVAGRMAGATAOGLIRAGKVEAQHILASAPTRNLCHFQALGCRTH 60
 DB 1 MAAAEPSPRVGFVAGRMAGATAOGLIRAGKVEAQHILASAPTRNLCHFQALGCRTH 60
 QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEAPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120

Db 61 SNOEVLOSCLLVIFATKPHVLPVLAEPVWVTTTEHILVSVAGMSLSTLEELLPPNTRV 120
QY 121 LRVLPNLPVVOEGAIIVMARGHRVGSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGSG 180
Db 121 LRVLPNLPVVOEGAIIVMARGHRVGSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGSG 180
QY 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALEOGGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALEOGGLRAATMSAVEAATCRAKELSRK 274

RESULT 3
AAG66956
ID AAG66956 standard; Protein; 274 AA.

AC AAG66956;
XX 22-OCT-2001 (first entry)
DT Human dihydropyrrrole-5-carboxylate reductase 30 polypeptide.
DE Human dihydropyrrrole-5-carboxylate reductase 30 polypeptide.
XX Human; dihydropyrrrole-5-carboxylate reductase 30; cancer; cytostatic;
KW human immunodeficiency virus; HIV; infection; immunological disease;
KW inflammatory disease.
XX Homo sapiens.
OS CNL298002-A.
PN 06-JUN-2001.
PD 24-NOV-1999; 99CN-0124090.
PF 24-NOV-1999; 99CN-0124090.
PR (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
PA Mao Y, Xie Y;
PI WPI: 2001-489680/54.
DR N-PSDB; AAH77597.
XX Human dihydropyrrrole-5-carboxylate reductase 30 as one new kind of
PT polypeptide and polynucleotides encoding this polypeptide -
XX Claim 1; Page 20-21 (disclosure); 26pp; Chinese.
PS The invention relates to a novel polypeptide, human
CC dihydropyrrrole-5-carboxylate reductase 30, polynucleotides encoding
CC this polypeptide and a DNA recombination process to produce the
CC polypeptide. The polypeptide is useful for treating various diseases,
CC such as malignant tumours, nosohaemia, HIV infection, immunological
CC diseases and inflammatory diseases. The invention also provides an
CC antibody against the polypeptide. The present sequence is the
CC polypeptide of the invention.

XX Sequence 274 AA;
SQ Query Match 99.88; Score 1383; DB 22; Length 274;
Best Local Similarity 99.68; Pred. No. 1.1e-130;
Matches 273; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAEPPRRVGFVAGRMAGATAOGLIRAGKVEAOHILASPTDRNLCHFQALGCRTH 60
Db 1 MAAEPPRRVGFVAGRMAGATAOGLIRAGKVEAOHILASPTDRNLCHFQALGCRTH 60
QY 61 SNOEVLOSCLLVIFATKPHVLPVLAEPVWVTTTEHILVSVAGMSLSTLEELLPPNTRV 120
|||||

Db 61 SNOEVLOSCLLVIFATKPHVLPVLAEPVWVTTTEHILVSVAGMSLSTLEELLPPNTRV 120
QY 121 LRVLPNLPVVOEGAIIVMARGHRVGSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGSG 180
Db 121 LRVLPNLPVVOEGAIIVMARGHRVGSSETKLLQHLLEACGRCEEVPEAYVDIHTGLSGSG 180
QY 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
Db 181 VAFVCAFSALAEAGAVKMGMPSSLAHRIAQTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
QY 241 TIYGLHALEOGGLRAATMSAVEAATCRAKELSRK 274
Db 241 TIYGLHALEOGGLRAATMSAVEAATCRAKELSRK 274
RESULT 4
ABP41278
ID ABP41278 standard; Protein; 343 AA.
XX AC ABP41278;
XX 22-AUG-2002 (first entry)
DT Human ovarian antigen HOGDC67, SEQ ID NO:2410.
DE Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory system disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antinflammatory; gynaecological; reproductive; chromosome 17.
XX Homo sapiens.
OS WO200200677-A1.
PN 03-JAN-2002.
PD 07-JUN-2001; 2001WO-US18569.
PF 07-JUN-2000; 2000US-209467P.
PR (HUMA-) HUMAN GENOME SCI INC.
PA Birse CE, Rosen CA;
PI WPI: 2002-147878/19.
DR N-PSDB; ABQ54355.
DR Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX Claim 11; SEQ ID No 2410; 2922pp; English.
XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic

AAE12785
ID AAE12785 standard; Protein; 315 AA.
AC
XX AAE12785;
XX
DT 15-JAN-2002 (first entry)
XX
DE Human delta 1-pyrroline-5-carboxylate reductase (P5CR).
XX
XX Human; delta 1-pyrroline-5-carboxylate reductase; P5CR; cystic fibrosis;
KW osteoporosis; neuronal disorder; gene therapy; akathisia; drug screening;
KW actinic keratosis; Alzheimer's disease; amyotrophic lateral sclerosis;
KW connective tissue disorder; myocardial fibrosis; cell proliferation;
KW arteriosclerosis bursitis; cancer; amnesia; neuroprotective; cytostatic;
KW antiarteriosclerotic; osteopathic; cardiant.
XX
OS Homo sapiens.
XX
XX US6268192-B1.
PN
XX
XX 31-JUL-2001.
PD
XX
XX 05-MAY-2000; 2000US-0565910.
PF
XX
XX 18-JUN-1998; 98US-0099676.
PR
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Hillman JL, Corley NC, Baughn MR;
PI
XX
XX WPI; 2001-647056/74.
DR
XX
XX New delta 1-pyrroline-5-carboxylate reductase polypeptides and
PT polynucleotides, useful for diagnosing, treating and preventing
PT neuronal disorders, connective tissue disorders or disorders of cell
PT proliferation
XX
PS Disclosure; Fig 2; 32pp; English.
XX
XX The present invention relates to delta 1-pyrroline-5-carboxylate
CC reductase (P5CR) polypeptides and polynucleotides. The P5CR gene is used
CC in gene therapy. The P5CR nucleic acid and amino acid sequences are
CC useful in the diagnosis, treatment and prevention of neuronal disorders
CC (e.g. akathisia, Alzheimer's disease, amnesia, or amyotrophic lateral
CC sclerosis), connective tissue disorders (e.g. cystic fibrosis, myocardial
CC fibrosis, or osteoporosis) and disorders of cell proliferation (e.g.
CC actinic keratosis, arteriosclerosis bursitis, cancers). P5CR may also be
CC used for screening libraries of compounds in various drug screening
CC techniques. The present sequence is human delta 1-pyrroline-5-carboxylate
CC reductase (P5CR).
XX
XX Sequence 315 AA;
SQ
Query Match 40.4%; Score 560.5; DB 22; Length 315;
Best Local Similarity 46.1%; Pred. No. 8e-48;
Matches 123; Conservative 46; Mismatches 87; Indels 11; Gaps 5;
QY 11 VGFVAGRMAGAIAGLIRAGKVEAQHILASAPTRNLCHFOAL---GCRTHSNQEVLO 67
Db 3 VGFIGAGQL--AFAGKGTAAAGVLAHAKIMASSP-DMDLATVSALRKMGVLTLPNKETVQ 59
QY 68 SCLLVIFATKPHVLPVLAIEAVPVVTTTEHILVSVAGVSLSTLEELLP---PNTNRLVRL 124
Db 60 HSDVFLAVKPHIIPFILDIGADIEDRHVVSCAAGVTISSIEKLSAFRAPVIRCM 119
QY 125 PNLPCVQGEAIVMARGHVGSSETKLLOHILLEACGCEVEPPEAYVDIHGLSSGVAFV 184
Db 120 TNPVVRGATVATGTGTHAQVDEGLMEQLLSTVGCTEVEEDLIDAVTGLSSGPAY- 178
QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYG 244
Db 179 -AFTALDADGGVKMGLPRRLAVRLGAQALLGAARKMLLHSEQHPQLKDNVSSPGGATIIA 237

QY 245 LHALEQGGRLRAATMSAVEAATCRAKEL 271
Db 238 LHVLESGGFRSLLINAVEASCIRTREL 264
RESULT 9
AAB74779
ID AAB74779 standard; Protein; 320 AA.
XX
XX AAB74779;
AC
XX 06-JUN-2001 (first entry)
DT
XX Human Py-CR protein SEQ ID NO:4.
DE
XX Human; pyrroline 5'-carboxylate reductase; Py-CR; P5CR.
KW
XX Homo sapiens.
OS
XX CN1274728-A.
PN
XX
XX 29-NOV-2000.
PD
XX
XX 25-MAY-1999; 99CN-0107071.
PF
XX
XX 25-MAY-1999; 99CN-0107071.
PR
XX (UYFU-) UNIV FUDAN.
PA
XX
XX Yu L, Fu O, Zhang H;
PI
XX
XX WPI; 2001-211749/22.
DR
XX N-PSDB; AAF81847.
DR
XX
XX New human protein and its code sequence, preparation and application -
PT
XX
XX Claim 4; Page 16; 20pp; Chinese.
PS
XX
XX The present invention describes a human protein designated Py-CR, which
CC is homologous to the human pyrroline-5'-carboxylate reductase (P5CR)
CC (EC1.5.1.2). The present invention also describes methods for the
CC application and production process of the py-CR polynucleotide and
CC protein sequences. The present sequence represents the human Py-CR
CC protein as given in the present invention.
XX
XX Sequence 320 AA;
SQ
Query Match 40.4%; Score 560.5; DB 22; Length 320;
Best Local Similarity 45.7%; Pred. No. 8.2e-48;
Matches 122; Conservative 48; Mismatches 90; Indels 7; Gaps 3;
QY 11 VGFVAGRMAGAIAGLIRAGKVEAQHILASAPTRNLCHFOAL---GCRTHSNQEVLO 67
Db 3 VGFIGAGQLANALARGPTAAGILSAHKIIASSP-EMNLPTVSALRKMGVNLTRSNKETVK 61
QY 68 SCLLVIFATKPHVLPVLAIEAVPVVTTTEHILVSVAGVSLSTLEELLP---PNTNRLVRL 124
Db 62 HSDVFLAVKPHIIPFILDIGADIEADQARHIVVSCAAGVTISSIEKLMFAQPAKVIKCM 121
QY 125 PNLPCVQGEAIVMARGHVGSSETKLLOHILLEACGCEVEPPEAYVDIHGLSSGVAFV 184
Db 122 TNPVVRGATVATGTGTHALVEGQLEQLMSSVGCTEVEEDLIDAVTGLSSGPAYA 181
QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAATLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYG 244
Db 182 FMAILDADADGGVKMGLPRRLAIQGAQALLGAARKMLLHSEQHPQLKDNVSSPGGATIIA 241
QY 245 LHALEQGGRLRAATMSAVEAATCRAKEL 271
Db 242 LHVLESGGFRSLLINAVEASCIRTREL 268
RESULT 10


```

Query Match      38.8%; Score 537.5; DB 21; Length 314;
Best Local Similarity 45.3%; Pred. No. 1.6e-43;
Matches 121; Conservative 45; Mismatches 88; Indels 13; Gaps 5;

Qy 11 VGFVCGAGMAGAIQAQLIRACKVEQAQIILASAPTRDLNLCHFQAL---GCRTHHSNQEVLQ 67
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3 VGFVIGAGQ----LAVRTAAGILSAHKILIASSP-EMNLPTVSALRKMVGVNLTIRSNKETIVK 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


us-09-806-536a-14.rag

Mon Jul 28 08:07:02 2003

PR	09-MAR-1999;	99US-0123548.	PR	15-JUL-1999;	99US-0144005.
PR	23-MAR-1999;	99US-0125788.	PR	16-JUL-1999;	99US-0144085.
PR	25-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0129845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0131449.	PR	20-JUL-1999;	99US-0144884.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0144814.
PR	04-MAY-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145086.
PR	04-MAY-1999;	99US-0132484.	PR	21-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132863.	PR	22-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	99US-0134256.	PR	22-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134221.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	99US-0134941.	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	99US-0135124.	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146388.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0146389.
PR	27-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	99US-0137222.	PR	04-AUG-1999;	99US-0147302.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147192.
PR	04-JUN-1999;	99US-0137502.	PR	05-AUG-1999;	99US-0147260.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147303.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139115.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	17-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139763.	PR	26-AUG-1999;	99US-0151065.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151066.
PR	22-JUN-1999;	99US-0139899.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0140353.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140354.	PR	31-AUG-1999;	99US-0151338.
PR	24-JUN-1999;	99US-0140695.	PR	01-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	99US-0140823.	PR	07-SEP-1999;	99US-0152363.
PR	30-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153070.
PR	01-JUL-1999;	99US-0141842.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0142154.	PR	16-SEP-1999;	99US-0154018.
PR	02-JUL-1999;	99US-0142055.	PR	20-SEP-1999;	99US-0154039.
PR	06-JUL-1999;	99US-0142390.	PR	22-SEP-1999;	99US-0154779.
PR	08-JUL-1999;	99US-0142803.	PR	23-SEP-1999;	99US-0155486.
PR	09-JUL-1999;	99US-0142920.	PR	24-SEP-1999;	99US-0155659.
PR	12-JUL-1999;	99US-0142577.	PR	28-SEP-1999;	99US-0156458.
PR	13-JUL-1999;	99US-0143542.	PR	29-SEP-1999;	99US-0156596.
PR	14-JUL-1999;	99US-0143624.	PR	04-OCT-1999;	99US-0157117.
			PR	05-OCT-1999;	99US-0157753.

WO200177335-A2.
 18-OCT-2001.
 11-APR-2001; 2001WO-FR01118.
 11-APR-2000; 2000FR-0004629.
 (INSP) INST PASTEUR.
 Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 Daniels J, Goebel W, Kref J, Kuhn M, Ng E, Vazquez-Boland JA;
 Dominguez-Bernal G, Garrido-Garcia P, Tierraz-Martinez A, Amend A;
 Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 Rose M, Voss H;
 WPI; 2002-010914/01.
 Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
 and prevention of *Listeria* and related bacterial infections, and
 related polypeptides
 Claim 6; SEQ ID No 132; 192pp; French.
 The present invention relates to the genome sequence of *Listeria*
monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 it are useful for selecting probes and primers for detecting genes in *L.*
monocytogenes and related organisms, and for studying genetic
 polymorphisms and other genomes. The present sequence is a protein
 encoded by the genome sequence of the present invention. Proteins
 expressed from the genome sequence are useful for raising specific
 antibodies, identification of *L. monocytogenes* and related organisms, and
 for biosynthesis and biodegradation, especially biosynthesis of vitamin
 B12. The genome sequence and proteins encoded by it are also useful for
 selecting compounds that regulate gene expression and cell replication
 and modulate *L. monocytogenes*-related diseases. In addition, the genome
 sequence and proteins encoded by it are useful in pharmaceutical and
 vaccines compositions for the treatment or prevention of infections by *L.*
monocytogenes and related organisms.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.
 Sequence 266 AA;
 Query Match 33.8%; Score 459; DB 23; Length 266;
 Best Local Similarity 38.8%; Pred. No. 1e-38;
 Matches 101; Conservative 55; Mismatches 96; Indels 8; Gaps 2
 QY 10 RVGFVAGRMAGAIAGQLIRAGKVEAQHILASAPTDRNLCHFQAL-----GCRTHSNQE 184
 Db 3 RIGFVGAGMGWTAMIRGLAQANLVKREIIVGG--RNLEKLKPLEAEFTGLQITTDTEK 59
 QY 65 VLQSCLLVIFATKPHVLPAVLAEPVYVTTTHILVSVAGVSLSTLEELLPPNTRVLRL 124
 Db 60 LVEQADIILAVKPTTPEITLTSVKREKLTDPKLIISVAAAGVTIQDLSELTSAKTKIVRM 119
 QY 125 PNLPCVQVEGALVMARGHVSSESTKLLQHLLEACGCEVEPYAVDVIHTGLSGGVAFV 184
 Db 120 PNTPALVGEAMSSVSPNTNVTSEELKEVTAFTFSFGAEVYVSENLMDAVIGVSGSPAYV 179
 QY 185 CAFSEALAGVAKMGMPSSLAHRITANOVTLLGTAKWLLHEGHPALRSDVCTPGGTTIYG 244
 Db 180 YMFIEIADGAGVLSGMPKDXKAYKFAQAQAVGAATVLETGHPGKGLKDMWTSPEGTTIEA 239
 QY 245 LHAEQGGDLRAATMSAVEAA 264
 Db 240 VKSLNDGFRSAVINAVQAA 259

QY 256 ATMSAVEATCRAKEL 271
Db 189 LLINAVEASCIRTREL 204

Search completed: July 21, 2003, 09:53:14
Job time : 73 secs

RESULT 15
AAB56640
ID AAB56640 standard; Protein; 255 AA.
XX
AC AAB56640;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1218.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytosolic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 200WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR N-PSDB; AAF15843.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 11; Page 1636-1637; 2338pp; English.
XX
CC AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
SQ Sequence 255 AA;
Query Match 33.7%; Score 467.5; DB 21; Length 255;
Best Local Similarity 48.5%; Pred. No. 1.4e-38;
Matches 95; Conservative 34; Mismatches 64; Indels 3; Gaps 1;
QY 79 HVLPAVLAEPVPTTEHILVSVAGVSLTLELLP---PNTRYLRVLPNLPVVOEGA 135
Db 9 HIIFPILDEIGADIEDRHIVVCAAGVTISSIEKKLSAFRPAFRPVRMTNTPVVVREGA 68
QY 136 IVMARGHVGSSTKLLQHLLEACGCEEPYAYVDIHTGLSGGVAFVCFSEALAEGA 195
Db 69 TVYATGTHAQVEDGRLEMEQLLSVGFCTEVEDLIDAVTGLSSGSPAYAFALDALDGG 128
QY 196 VKMGMPSSLAHRTAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGLHLEOGGLRA 255
Db 129 VKMGLPRRLAVRLGAQALLGAARKMLLSHQHPGQKLDNVSPGGATIHALHVLSEGGFRS 188

(INCY-) INCYTE PHARM INC.

Lal P, Guegler KJ, Gorgone GA, Corley NC, Baughn MR, Hillman JL, Bandman O, Azimzai Y, Au-Young J, Yue H, Yang J; Tang YT; Lu DAM;

WPI; 2000-303785/26.
P-PSDB; AAY92517.

Purified polypeptide for treating or preventing disorders associated with decreased expression or activity of oxidoreductase molecules

Claim 9; Page 96; 97pp; English.

AAA09375-89 encode oxidoreductases, designated OXRE-1 to -15. OXRE-14 has identity with pyrroline-5-carboxylate-reductase. The polypeptides are useful for treating or preventing a disorder associated with decreased expression or activity of OXRE. Antagonists of OXRE are useful for treating or preventing a disorder associated with increased expression or activity of OXRE. The disorders include cell proliferative disorders (cirrhosis, hepatitis), cancer (leukemia, melanoma), hypopituitarism and hyperpituitarism, hypothyroidism and hyperthyroidism, metabolic disorders (Addison's disease, cystic fibrosis), reproductive disorders (infertility, ovulatory defects), neurological disorders (Alzheimer's disease, Parkinson's disease, multiple sclerosis), mental disorders (anxiety, schizophrenia), autoimmune/inflammatory disorders (acquired immunodeficiency syndrome (AIDS), asthma, osteoarthritis), and viral infections. The polynucleotides may be used in Southern or Northern analysis. The chain reaction (PCR), or in enzyme-linked immunosorbent assays (ELISA).

Sequence 1338 BP; 225 A; 423 C; 440 G; 250 T; 0 other;

Query Match	100.0%	Score 1338;	DB 21;	Length 1338;
Best Local Similarity	100.0%	Pred. No. 8e-289;		
Matches 1338; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	GGTGAGCCAGCTCTGTCGGAGGCACAAAGATGCGAGCTCGGAGCGCGTCTCCGCGGCGCG	60
Db	1	GGTGAGCCAGCTCTGTCGGAGGCACAAAGATGCGAGCTCGGAGCGCGTCTCCGCGGCGCG	60
QY	61	TGGGCTTCGTGGCGCGCGGCCGATGCGGGGGGCCATCGCGAGGGCCTTCATCAGACGAG	120
Db	61	TGGGCTTCGTGGCGCGCGGCCGATGCGGGGGGCCATCGCGAGGGCCTTCATCAGACGAG	120
QY	121	GAAGAAGTGGAGGCTCAGCACAATACTGGCCAGTGCACCAACAGACAGGAACCTATGTCACT	180
Db	121	GAAGAAGTGGAGGCTCAGCACAATACTGGCCAGTGCACCAACAGACAGGAACCTATGTCACT	180
QY	181	TTCAAGCTCTGGGTTGGCGGACCACGCACTCCAACCAGGAGCTGCTGCAGAGCTGCCTGCG	240
Db	181	TTCAAGCTCTGGGTTGGCGGACCACGCACTCCAACCAGGAGCTGCTGCAGAGCTGCCTGCG	240
QY	241	TCGTCACATTTGGCCCAAGCCCTCATGTGGTGGCAGCTGCTCTGGCAGAGGTGGCTCCTG	300
Db	241	TCGTCACATTTGGCCCAAGCCCTCATGTGGTGGCAGCTGCTCTGGCAGAGGTGGCTCCTG	300
QY	301	TGGTCACCACATGAACACATTTGGTGTCCGTGGCTGCTGGGGTGCTCTGAGACACCTTGG	360
Db	301	TGGTCACCACATGAACACATTTGGTGTCCGTGGCTGCTGGGGTGCTCTGAGACACCTTGG	360
QY	361	AGGAGCTTGCTGCCCCCAACACACAGGGTGCTCGGGGTCTTGCCCCAACTGCCCTGTGTGG	420
Db	361	AGGAGCTTGCTGCCCCCAACACACAGGGTGCTCGGGGTCTTGCCCCAACTGCCCTGTGTGG	420
QY	421	TCCAGGAAGGGGCGCATAGTGATGGCGCGGGCGCGCAGCTGGGGAGCAGCAGACCAAGC	480
Db	421	TCCAGGAAGGGGCGCATAGTGATGGCGCGGGCGCGCAGCTGGGGAGCAGCAGACCAAGC	480
QY	481	TCCTGCAGCATCTGCTTGGAGGCGCTGTGGCGGTGTGAGGAGGTGCCTGAAGCCTTACGTCG	540
Db	481	TCCTGCAGCATCTGCTTGGAGGCGCTGTGGCGGTGTGAGGAGGTGCCTGAAGCCTTACGTCG	540

Qy	541	ACATCCACACTGGCCCTCAGTGGCAGTGGCGTGGCCCTTCGTGTGCATCTCTCCGAGGCC	600
Db	541		
Qy	541	ACATCCACACTGGCCCTCAGTGGCAGTGGCGTGGCCCTTCGTGTGTGCATCTCTCCGAGGCC	600
Db	541		
Qy	601	TGGCTGAAGGAGCCGTCAAGATGGCATGCCACAGAGCCTGGCCCAACGCATCGCTGGCC	660
Db	601		
Qy	601	TGGCTGAAGGAGCCGTCAAGATGGCATGCCACAGAGCCTGGCCCAACGCATCGCTGGCC	660
Db	601		
Qy	661	AGACCCCTGCTGGGACGCCAAGATGCTGCTGCAGAGGGCCAAACACCCAGGCCCAAGCTGC	720
Db	661		
Qy	661	AGACCCCTGCTGGGACGCCAAGATGCTGCTGCAGAGGGCCAAACACCCAGGCCCAAGCTGC	720
Db	661		
Qy	721	GCTCAGAGCTGTGCACCCCGGGTGGCACACACATCTATGGACTCCACGCCCTTGGAGCAGG	780
Db	721		
Qy	721	GCTCAGAGCTGTGCACCCCGGGTGGCACACACATCTATGGACTCCACGCCCTTGGAGCAGG	780
Db	721		
Qy	781	CGGGCTCGAGAGCACCACATGAGCGCGTGGAGGCTGCCACCTGCGCGGCCCAAGGAGC	840
Db	781		
Qy	781	CGGGCTCGAGAGCACCACATGAGCGCGTGGAGGCTGCCACCTGCGCGGCCCAAGGAGC	840
Db	781		
Qy	841	TCAGCAGAAAGTAGGCTGGGCTCTGGCCATCCTTTCCTGCCCTCTGTGCCCTTGCCTCTCC	900
Db	841		
Qy	841	TCAGCAGAAAGTAGGCTGGGCTCTGGCCATCCTTTCCTGCCCTCTGTGCCCTTGCCTCTCC	900
Db	841		
Qy	901	CTGTGTCCCTTCGCCCTGAGGACTGGGCTCCCTCCCTCGATGAGGGTCTCTACTGTC	960
Db	901		
Qy	901	CTGTGTCCCTTCGCCCTGAGGACTGGGCTCCCTCCCTCGATGAGGGTCTCTACTGTC	960
Db	901		
Qy	961	TCCTTCTCCCTTCACAGGGAAATGAGGGGGCAGGACTTGGAGGTTTCCAGCAGGCGG	1020
Db	961		
Qy	961	TCCTTCTCCCTTCACAGGGAAATGAGGGGGCAGGACTTGGAGGTTTCCAGCAGGCGG	1020
Db	961		
Qy	1021	GGGAGCCCCGACCACTGGGGACACTCTCTCCCTCCCAAGTAGCAGAAAGGCCCTGGTGG	1080
Db	1021		
Qy	1021	GGGAGCCCCGACCACTGGGGACACTCTCTCCCTCCCAAGTAGCAGAAAGGCCCTGGTGG	1080
Db	1021		
Qy	1081	TGGCTCTGCCCTTGCTGCACTGAGCCCCACCTTCTCTCAACATTTGTTCTGAGGGGGCCA	1140
Db	1081		
Qy	1081	TGGCTCTGCCCTTGCTGCACTGAGCCCCACCTTCTCTCAACATTTGTTCTGAGGGGGCCA	1140
Db	1081		
Qy	1141	AGAGATGGCGTCTTGGTCATTTGGCCCGCATGGTTGGGCGAGTTGGTTCAGGCCCATGAACAG	1200
Db	1141		
Qy	1141	AGAGATGGCGTCTTGGTCATTTGGCCCGCATGGTTGGGCGAGTTGGTTCAGGCCCATGAACAG	1200
Db	1141		
Qy	1201	AACTTAGCGTTAACAGGACCGGCTGGCCCAATGCCGTGTGGAGCTGGAGCTTGGCTTTG	1260
Db	1201		
Qy	1201	AACTTAGCGTTAACAGGACCGGCTGGCCCAATGCCGTGTGGAGCTGGAGCTTGGCTTTG	1260
Db	1201		
Qy	1261	GCCTTCCAAAGTGGGCTCGTGCAGCTACAGCCAGGCCGGCTGCCTCACTCTCAGCTCTAGGG	1320
Db	1261		
Qy	1261	GCCTTCCAAAGTGGGCTCGTGCAGCTACAGCCAGGCCGGCTGCCTCACTCTCAGCTCTAGGG	1320
Db	1261		
Qy	1321	GGCAGCAGCATATGGGGT	1338
Db	1321		
Qy	1321	GGCAGCAGCATATGGGGT	1338
Db	1321		

RESULT 2
AAH18288
ID AAH18288 standard; cDNA; 2331 BP.

1

PD	07-FEB-2001.	QY	315	CACATCTGTGTCCTGGTGTCTCTGAGCACCTGGAGAGCTGCTGCC	374
XX					
PF	28-JUL-2000; 2000BP-0116126.	Db	312	CACATCTGTGTCCTGGTGTCTCTGAGCACCTGGAGAGCTGCTGCC	371
XX					
PR	29-JUL-1999; 99JP-0248036.	QY	375	CCAAACACACGGGTGCTGCCGGTCTTCCCAACCTGTCCTGTCGAGGAGGAGG	434
PR	27-AUG-1999; 99JP-0300253.				
PR	11-JAN-2000; 2000JP-0118776.	Db	372	CCAAACACACGGGTGCTGCCGGTCTTCCCAACCTGTCCTGTCGAGGAGGAGG	431
PR	02-MAY-2000; 2000JP-0183767.				
PR	09-JUN-2000; 2000JP-0241899.	QY	435	ATAGTATGTCGGCGGGCGCCACGTGGGAGCAGACCAAGCTCTGACGATCTG	494
XX					
PA	(HELI-) HELIX RES INST.	Db	432	ATAGTATGTCGGCGGGCGCCACGTGGGAGCAGACCAAGCTCTGACGATCTG	491
XX					
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	QY	495	CTGAGGCTGTGGCGGTGTGAGAGTGCTTGAAGCTTACGTCGACATCCACATG	554
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;				
PI	WPI; 2001-318749/34.	Db	492	CTGAGGCTGTGGCGGTGTGAGAGTGCTTGAAGCTTACGTCGACATCCACATG	551
XX					
XX	Primer sets for synthesizing polynucleotides, particularly the 5602	QY	555	CTCAGTGCACTGGCGGTGGCCCTTCTGTGTGCATTCCTCCGAGGCCCTGGCTGA	614.
PT	full-length cDNAs defined in the specification, and for the detection				
PT	and/or diagnosis of the abnormality of the proteins encoded by the	Db	552	CTCAGTGCACTGGCGGTGGCCCTTCTGTGTGCATTCCTCCGAGGCCCTGGCTGA	611
PT	full-length cDNAs -				
XX		QY	615	GTCAGATGGCATGCTGCCAGCAGCTGCCCCACCGCATCTGCTGCCAGACCTGCT	674
XX					
PS	Claim 8; SEQ ID 18268; 2537pp + CD ROM; English.	Db	612	GTCAGATGGCATGCTGCCAGCAGCTGCCCCACCGCATCTGCTGCCAGACCTGCT	671
XX					
XX	The present invention describes primer sets for synthesizing 5602	QY	675	ACGCCAAGATGCTGCTGCACGAGGGCCACACAGCCAGCTGCTCAGAGCTGCTG	734
CC	full-length cDNAs defined in the specification. Where a primer set				
CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary	Db	672	ACGCCAAGATGCTGCTGCACGAGGGCCACACAGCCAGCTGCTCAGAGCTGCTG	731
CC	to the complementary strand of a polynucleotide which comprises one of				
CC	the 5602 nucleotide sequences defined in the specification, where the	QY	735	ACCCGGGTGGCACCACCATCTATGGACTCCACGCCCTGGAGCAGGGGGCTGGG	794
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination				
CC	of an oligonucleotide comprising a sequence complementary to the	Db	732	ACCCGGGTGGCACCACCATCTATGGACTCCACGCCCTGGAGCAGGGGGCTGGG	791
CC	complementary strand of a polynucleotide which comprises a 5'-end				
CC	sequence and an oligonucleotide comprising a sequence complementary to a	QY	795	GCCACCATGAGCGCGTGGAGCTGCCACCTGCCGGGCCAAAGAGCTCAGCAGAA	854
CC	polynucleotide which comprises a 3'-end sequence, where the				
CC	oligonucleotide comprises at least 15 nucleotides and the combination of	Db	792	GCCACCATGAGCGCGTGGAGCTGCCACCTGCCGGGCCAAAGAGCTCAGCAGAA	851
CC	the 5'-end sequence/3'-end sequence is selected from those defined in				
CC	the specification. The primer sets can be used in antisense therapy and	QY	855	GCTGGGCTCTGGCCATCTCTTCTGCTGCTGTGCCCTCTGCTCCTCTGCTGCTCC	914
CC	in gene therapy. The primers are useful for synthesizing polynucleotides,				
CC	particularly full-length cDNAs. The primers are also useful for the	Db	852	GCTGGGCTCTGGCCATCTCTTCTGCTGCTGTGCCCTCTGCTCCTCTGCTGCTCC	911
CC	detection and/or diagnosis of the abnormality of the proteins encoded by				
CC	the full-length cDNAs. The primers allow obtaining of the full-length	QY	915	CTGAGGACTGGGCTCCCTCCCTCTGATGAGGCTCTCTCTCTCTCTCTCTCTCT	974
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and				
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to	Db	912	CTGAGGACTGGGCTCCCTCCCTCTGATGAGGCTCTCTCTCTCTCTCTCTCTCT	971
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632				
CC	represent oligonucleotides, all of which are used in the exemplification	QY	975	CACAGGAAATGCAGGGGCGCAGGACTTGGAGGTTCCAGCAGGGGGGAGCCCGAC	1034
CC	of the present invention.				
XX		Db	972	CACAGGAAATGCAGGGGCGCAGGACTTGGAGGTTCCAGCAGGGGGGAGCCCGAC	1031
SQ	Sequence 2331 BP; 398 A; 753 C; 748 G; 432 T; 0 other;				
	Query Match 97.7%; Score 1307; DB 22; Length 2331;	QY	1035	GTGGGACACTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT	1094
	Best Local Similarity 99.2%; Pred. No. 7.4e-282;				
	Matches 1313; Conservative 0; Mismatches 10; Indels 0; Gaps 0;	Db	1032	GTGGGACACTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT	1091
QY	15 GTCGAGGCAACAAGATGCGAGCTGCGGAGCCCTCTCCGCGGCGCTGGCTTCG	QY	1095	GCTCAGTGCAGCCACCTTGTGCAACATTTGTTCTGAGGGGCCCAAGAGATGGG	1154
Db	12 GTCGAGGCAACAAGATGCGAGCTGCGGAGCCCTCTCCGCGGCGCTGGCTTCG	Db	1092	GCTCAGTGCAGCCACCTTGTGCAACATTTGTTCTGAGGGGCCCAAGAGATGGG	1151
QY	75 GCGGGCGCATGCGGGGCGCATCGCGAGGGCTCATCAGACAGGAAAGTGAAG	QY	1155	GCTCATTTGCGCGCATGTTGGGCGAGTTGGTTAGGCCCATGAACAGAACTTAC	1214
Db	72 GCGGGCGCATGCGGGGCGCATCGCGAGGGCTCATCAGACAGGAAAGTGAAG	Db	1152	GCTCATTTGCGCGCATGTTGGGCGAGTTGGTTAGGCCCATGAACAGAACTTAC	1211
QY	135 CAGCACATACTGGCCAGTGCACCAACAGACAGAACCTATGTCACTTTCAAG	QY	1215	GGCAGGGCTGGCCCAATGCTGGTCTGGAGCTGGAGCTTGGCTTTTCCCAAGT	1274
Db	132 CAGCACATACTGGCCAGTGCACCAACAGACAGAACCTATGTCACTTTCAAG	Db	1212	GGCAGGGCTGGCCCAATGCTGGTCTGGAGCTGGAGCTTGGCTTTTCCAGAGT	1271
QY	195 TGCGGACACAGCACTCCACACAGGAGGTGCTGCAGAGCTGCTGCTCATCTT	QY	1275	CTCTGTCAGCTACAGCCAGGGCGGCTGCTCATCTCAGCTCTAGGGGGCAGCAG	1334
Db	192 TGCGGACACAGCACTCCACACAGGAGGTGCTACAGAGCTGCTGCTCATCTT	Db	1272	TCGTCAGCTACAGCCAGGGCGGCTGCTCATCTCAGCTCTAGGGGGCAGCAG	1331
QY	255 ACCAAGCCTCATGTGCTGCCAGTGTCTGCGCAGAGGTGGCTCTCTGTGGT	QY	1335	GGG 1337	
Db	252 ACCAAGCCTCATGTGCTGCCAGTGTCTGCGCAGAGGTGGCTCTCTGTGGT	Db	1332	GGG 1334	

ID	AAH14370 standard; cDNA; 2615 BP.
XX	AAH14370;
AC	26-JUN-2001 (first entry)
DT	Human cDNA sequence SEQ ID NO:11778.
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX	Homo sapiens.
OS	EP1074617-A2.
PX	07-FEB-2001.
PN	28-JUL-2000; 2000EP-0116126.
PD	29-JUL-1999; 99JP-0248036.
PF	27-AUG-1999; 99JP-0300253.
PR	11-MAY-2000; 2000JP-0118776.
PR	02-JAN-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	(HELI-) HELIX RES INST.
PA	Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J; Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T; WPI; 2001-318749/34.
XX	Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -
PT	Claim 8; SEQ ID 11778; 2537pp + CD ROM; English.
PS	The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primers can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.
CC	Sequence 2615 BP; 492 A; 831 C; 818 G; 474 T; 0 other;
SQ	Query Match Best Local Similarity 99.1%; Score 1291.8; DB 22; Length 2615; Matches 1299; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY	27 AAGATGCACCTCGGAGCGGTCTCCGGCGCGGTGGCTTCGTGGCGCGCGCATG 86 2615 AAGATGCACCTCGGAGCGGTCTCCGGCGCGGTGGCTTCGTGGCGCGCGCATG 2556
DB	87 GCGGGGCCATCGCGAGGCGCTCATCAGACGAGGAAGAAGTTGGAAGCTCAGCACATACTG 146
QY	

QY 1227 CCAATGCTGTCTGGAGCTGGAGCTTGCCTTTGGCTTCCAAAGTGGGCTCGTGACGCTA 1286
|||||
Db 1415 CCAATGCTGTCTGGAGCTGGAGCTTGCCTTTGGCTTCCAAAGTGGGCTCGTGACGCTA 1356
|||||
QY 1287 CAGCAGGCGGCTGCTCATCTCAGCTCTAGGGGCGACGACATATGGG 1337
|||||
Db 1355 CAGCAGGCGGCTGCTCATCTCAGCTCTAGGGGCGACGACATATGGG 1305
|||||

RESULT 5

AAH08500

ID AAH08500 standard; cDNA; 794 BP.

XX

AC AAH08500;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human cDNA clone (5'-primer) SEQ ID NO:5335.

XX

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

KW Homo sapiens.

XX

PN EP1074617-A2.

XX

XX 07-FEB-2001.

XX

XX 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.

XX

PR 27-AUG-1999; 99JP-0300253.

XX

PR 11-JAN-2000; 2000JP-0118776.

XX

PR 02-MAY-2000; 2000JP-0183767.

XX

PR 09-JUN-2000; 2000JP-0241899.

XX

XX (HELI-) HELIX RES INST.

XX

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX

XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

XX Claim 1; SEQ ID 5335; 2537pp + CD ROM; English.

XX

XX The present invention describes primer sets for synthesizing 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination

XX of an oligonucleotide comprising a sequence complementary to the

XX complementary strand of a polynucleotide which comprises a 5'-end

XX sequence and an oligonucleotide comprising a sequence complementary to a

XX polynucleotide which comprises a 3'-end sequence, where the

XX oligonucleotide comprises at least 15 nucleotides and the combination of

XX the 5'-end sequence/3'-end sequence is selected from those defined in

XX the specification. The primer sets can be used in antisense therapy and

XX in gene therapy. The primers are useful for synthesizing polynucleotides,

XX particularly full-length cDNAs. The primers are also useful for the

XX detection and/or diagnosis of the abnormality of the proteins encoded by

XX the full-length cDNAs. The primers allow obtaining of the full-length

XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

XX represent oligonucleotides, all of which are used in the exemplification

XX of the present invention.

XX

SQ Sequence 794 BP; 142 A; 248 C; 264 G; 133 T; 7 other;

Query Match 55.1%; Score 737.6; DB 22; Length 794;
Best Local Similarity 97.7%; Pred. No. 4.8e-155;
Matches 765; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

QY 15 GTCCGAGGCAACAAGATGGCAGCTGGGAGGCGCTCTCCCGCGGCGTGGGCTTCGTGGGC 74
|||
Db 12 GTCCGAGGCAACAAGATGGCAGCTGGGAGGCGCTCTCCCGCGGCGTGGGCTTCGTGGGC 71
|||
QY 75 GCGGCGCGCATGGCGGGGCGCATCGCGAGGGGCTCATCAGAGCAGGAGAAAGTGAAGCT 134
|||
Db 72 GCGGCGCGCATGGCGGGGCGCATCGCGAGGGGCTCATCAGAGCAGGAGAAAGTGAAGCT 131
|||
QY 135 CAGCACATCTGGCCAGTGCACCAACAGACAGAGAACCTATGTCACCTTTCAAGCTCTGGGT 194
|||
Db 132 CAGCACATCTGGCCAGTGCACCAACAGACAGAGAACCTATGTCACCTTTCAAGCTCTGGGT 191
|||
QY 195 TGCGGAGCACCGACCTCCAAACAGAGAGTCTCTCAGAGCTGCTGCTCATCTTTTGGC 254
|||
Db 192 TGCGGAGCACCGACCTCCAAACAGAGAGTCTCTCAGAGCTGCTGCTCATCTTTTGGC 251
|||
QY 255 ACCAAGCCTCATGTGCTGCCAGCTGTCTCGGACAGAGTGGCTCTGTGTGCTACCACTGAA 314
|||
Db 252 ACCAAGCCTCATGTGCTGCCAGCTGTCTCGGACAGAGTGGCTCTGTGTGCTACCACTGAA 311
|||
QY 315 CACATCTTGTGCTCGTGGTGTCTCTCAGCACCTGGAGGAGTGTGTGCTTGGC 374
|||
Db 312 CACATCTTGTGCTCGTGGTGTCTCTCAGCACCTGGAGGAGTGTGTGCTTGGC 371
|||
QY 375 CCAAACACACGGGTGTGCGGGTCTTGGCCAACTGCTGTGTGCTGAGGAGGAGGCGC 434
|||
Db 372 CCAAACACACGGGTGTGCGGGTCTTGGCCAACTGCTGTGTGCTGAGGAGGAGGCGC 431
|||
QY 435 ATAGTGATGCGCGGGGCGCCACGTCGCGGAGCAGCAGCAAGCTCC-TGCACATCT 493
|||
Db 432 ATAGTGATGCGCGGGGCGCCACGTCGCGGAGCAGCAGCAAGCTCTTGCACATCT 491
|||
QY 494 GCTGGAGGCTGTGGGCGGTGTGAGGAGTGTGCTGAAGCTCTAGTGGACATCCACACTGG 553
|||
Db 492 GCTGGAGGCTGTGGGCGGTGTGAGGAGTGTGCTGAAGCTCTAGTGGACATCCACACTGG 551
|||
QY 554 CCTCAGTGGCAGTGGCGTGGCTTCGTGTGTGCTGCTCCGAGGCGCTGGCTGAAGGAGC 613
|||
Db 552 CCTCAGTGGCAGTGGCGTGGCTTCGTGTGTGCTGCTCCGAGGCGCTGGCTGAAGGAGC 611
|||
QY 614 CGTCAAGATGGGATGCCAGCAGCCTGGCCACCACGATGCTGTGCCCCAGACCTCTGGG 673
|||
Db 612 CGTCAAGATGGGATGCCAGCAGCCTGGCCACCACGATGCTGTGCCCCAGACCTCTGGG 671
|||
QY 674 GACGGCCAAAGATGCTGCTGCACAGGCGCCAAACACCCAGCCAGCTGCGCTCAGAGCTGTG 733
|||
Db 672 GACGGCCAAAGATGCTGCTGCACAGGCGCCAAACACCCAGCCAGCTGNGCTCAGAGCTGTG 731
|||
QY 734 CA-CCCCGGGTGGCACCACCATCATATGGACTTCACGCCCTGGAGCAGGCGGCTGCGAG 792
|||
Db 732 CACCCCGGGTGGGACCAACCATCTATGGACTTCACGCCCTGGAGCAGGCGGCTGNNNA 791
|||
QY 793 CAG 795
|||
Db 792 CAG 794
|||

RESULT 6

AAK51999

ID AAK51999 standard; cDNA; 5055 BP.

XX

AC AAK51999;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polynucleotide SEQ ID NO 544.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
DR P-PSDB; AAM78866.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX Claim 1; Page 1961-1966; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX Sequence 5055 BP; 989 A; 1736 C; 1512 G; 818 T; 0 other;

Query Match 54.0%; Score 723; DB 22; Length 5055;
Best Local Similarity 95.9%; Pred. No. 1.3e-151;
Matches 750; Conservative 0; Mismatches 30; Indels 2; Gaps 1;

QY 45 CGCTCTCCGCGGGGTGGCTTCGTGGCGGGCGCGCATGGCGGGGGCCATCGCGCAG 104
DB 15 CGCTCTCCGCGGGGTGGCTTCGTGGCGGGCGGGCGCATGGC--GGGGCCATCGCGCAG 72
QY 105 GGCTCATCAGACGAGAAAGTGAAGCTCAGCACAATCTGGCCAGTGCACCAACAGAC 164
DB 73 GGCTCATCAGACGAGAAAGTGAAGCTCAGCACAATCTGGCCAGTGCACCAACAGAC 132
QY 165 AGAACCTATGCTATTTCAAGCTCTGGGTTCCGGACACGCATCTCAACAGGAGGTG 224
DB 133 AGGAACCTATGCTATTTCAAGCTCTGGGTTCCGGACACGCATCTCAACAGGAGGTG 192
QY 225 CTGCAGAGTGGCTGCTGCTCATCTTTGCCACCAAGCCTCATGCTGCCAGCTGCTG 284
DB 193 CTGCAGAGTGGCTGCTGCTCATCTTTGCCACCAAGCCTCATGCTGCCAGCTGCTG 252

QY 285 GCAGAGTGGCTCTCTGGTGCACACTGAACACATCTTGTGTCCGTGGCTGCTGGGGTG 344
DB 253 GCAGAGTGGCTCTCTGGTGCACACTGAACACATCTTGTGTCCGTGGCTGCTGGGGTG 312
QY 345 TCTGTAGCACCCCTGGAGGAGCTGCTGCCCCCCAAACACACGGGTGTGGGGTCTTGGCC 404
DB 313 TCTGTAGCACCCCTGGAGGAGCTGCTGCCCCCCAAACACACGGGTGTGGGGTCTTGGCC 372
QY 405 AACCTGCCCTGTGTGGTCCAGGAGGGGCCATAGTGATGGCGGGGGCCGACGTTGGG 464
DB 373 AACCTGCCCTGTGTGGTCCAGGAGGGGCCATAGTGATGGCGGGGGCCGACGTTGGG 432
QY 465 AGCAGCAGACCAAGCTCTGTCAGCATCTGTCGAGGCCCTGTGGCGGTGTGAGGAGGTG 524
DB 433 AGCAGCAGACCAAGCTCTGTCAGCATCTGTCGAGGCCCTGTGGCGGTGTGAGGAGGTG 492
QY 525 CCTGAAGCCTACGTCGACATCCACATGGGCTCAGTGGCAGTGGCGTTCGTGTGTGT 584
DB 493 CCTGAAGCCTACGTCGACATCCACATGGGCTCAGTGGCAGTGGCGTTCGTGTGTGT 552
QY 585 GCATTCTCCGAGGCCCTGGCTGAGGAGCCGTCAGATGGCATGCCACGACGCTGGCC 644
DB 553 GCATTCTCCGAGGCCCTGGCTGAGGAGCCGTCAGATGGCATGCCACGACGCTGGCC 612
QY 645 CACCGCATCGTGGCCAGACCCCTGCTGGGACGCCCAAGATGCTGTGCAGAGGGCCAA 704
DB 613 CACCGCATCGTGGCCAGACCCCTGCTGGGACGCCCAAGATGCTGTGCAGAGGGCCAA 672
QY 705 CACCCAGCCAGCTGCGCTCAGACGTGTGCACCCGGGTGGCACCACCATCTATGGA 764
DB 673 CACCCAGCCAGCTGCGCTCAGACGTGTGCACCCGGGTGGCACCACCATCTATGGA 732
QY 765 CACCCCTGGAGCAGCGGGGCTGCGAGCAGCCACCATGAGCGCGGTGGAGGCTGCCACC 824
DB 733 CACCCCTGGAGCAGCGGGGCTGATGCCAAGCCTGGGACGCTGATGCCAAGCCTGGG 792
QY 825 TG 826
DB 793 TG 794
RESULT 7
AAS42169
ID AAS42169 standard; DNA; 5647 BP.
XX AAS42169;
AC AAS42169;
XX 17-DEC-2001 (first entry)
XX Genomic sequence #485 encoding novel human enzyme polypeptide.
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
XX anti arthritic; nephrotropic; anticoagulant; ds.
OS Homo sapiens.
XX WO200155301-A2.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01239.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
PR XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465566/50.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Disclosure; SEQ ID No 2295; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the


```
Db 61 GCGGGGGCCATCGCGAGGCGCTCATCAGCAGGAAAGTGAAGCTCAGCACATACT 120
QY 146 GGCAGTGCACACACACAGGAACTATGCTCACTTCAAGCTCTGGTTGCCGACCAC 205
Db 121 GGCAGTGCACACACACAGGAACTATGCTCACTTCAAGCTCTGGTTGCCGACCAC 180
QY 206 GCACTCCAACAGAGAGTGCTGCAGAGCTGCCTGCTCGTCATCTTTGCCACCAAGCCTCA 265
Db 181 GCACTCCAACAGAGAGTGCTGCAGAGCTGCCTGCTCGTCATCTTTGCCACCAAGCCTCA 240
QY 266 TGTGCTGCCAGCTGCTCTGG - CAGAGGTGGCTCCTGTGFTCACTGAAACACATCTTGG 324
Db 241 TGTGCTGCCAGCTGCTCTGGACAGAGTGGCTCCTGTGFTCACTGAAACACATCTTGG 300
QY 325 TGTCCGTGGCTGCTGGGGTGTCTCTGAGCACCTGAGGAGCTGCTGCCCCCAACACAC 384
Db 301 TGTCCGTGGCTGCTGGGGTGTCTCTGAGCACCTGAGGAGCTGCTGCCCCCAACACAC 360
QY 385 GGTGCTGCGGGTCTTGCCCAACTGCCCTGTGTGTGTCAGGAAGGGGCATAG--TGAT 442
Db 361 GGTGCTGCGGGTCTTGCCCAACTGCCCTGTGTGTGTCAGGAAGGGGCATAGATGATG 420
QY 443 GCGCGGGGGCCGACGTGGGAGCAGCAGACCAAGCTCTGCGACGATCTGCTGGAGGC 502
Db 421 GCCTCGGGGGCCGACGTGGGAGCAGCAGACCAAGCTCTGCGACGATCTTCTGGATGC 480
QY 503 CTGTGGCGGTGTGAGGAGTGCTGAAAGCCTACGTGCGACATCCACACTGGCTCAGTGG 562
Db 481 CTGTGGCGGTGTGAGGAGTGCTGAAAGCCTACGTGCGACATCCACACTGGCTCAATGG 540
QY 563 CAGTGG-CGTGGCTTCGTGTGT 584
Db 541 CAGTGGCTGTGCGCTTCGNGT 563

RESULT 9
AAS41342
ID AAS41342 standard; cdNA; 270 BP.
XX
AC AAS41342;
XX
DT 17-DEC-2001 (first entry)
XX
DE cdNA encoding novel human enzyme polypeptide #558.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS Homo sapiens.
XX
PN WO20015301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226888.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
```


CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1889 BP; 364 A; 554 C; 589 G; 379 T; 3 other;

Query Match 16.8%; Score 224.8; DB 24; Length 1889;
Best Local Similarity 58.0%; Pred. NO. 1.2e-40;
Matches 466; Conservative 0; Mismatches 317; Indels 21; Gaps 3;

QY 56 GCGCGTGGCTTCGTGGCGGGCGGCATGCGGGGGCCATCGCGCAGGCGCTCATCAG 115
DB 1813 GAGCGTGGCTTCATCGCGCTGGCCAGCTGCTTTTCCCTGGCCAAAGGCTTCACAGC 1754
QY 116 AGCAGGAAAGTGAAGCTCAGCACAATCTGCGCAGTGCACCAACACAGACAGAACTATG 175
DB 1753 AGCAGGCTTCGTGGCTGCCACAGATATGGCTAGCTCC---CAGACATGGACCTGGC 1697
QY 176 TCACATTTCAAGCTC-----TGGGTTGCCGGACACAGCTCCAAACCAGGAGGTGCT 226
DB 1696 CACAGTTCCTCTCAGGAAGATGGGGGTGAAGTTGACAGCCCAACAAGAGACGGT 1637
QY 227 CGAGAGTGCCTGCTCGTCANCTTTGCCACCAAGCTCATGTGTCGACAGTGCCTGGC 286
DB 1636 GCAGCAGCATGATGTGCTCTTCTGGCTGTGAAGCCACACATCATCCCTTCATCCTGGA 1577
QY 287 AGAGGTGGCTCCTGTGGTCACACATGAACATCTTGGTGTCCGTGGCTGTGGGGTGTG 346
DB 1576 TGAATAGCGCGGACATGAGGACAGACACATTTGGTGTCTGCGCGGGCGGTGTCAC 1517
QY 347 TCTGAGCACCTTGGAGGAGTCTGTG-----CCCCCAACACACAGGCTGTGCGGGT 397
DB 1516 CATCAGCTCCATTGAGAAGAGCTGTACAGCGTTTCGGCCAGCCCGCAGGCTCATCGGTG 1457
QY 398 CTTGCCACCTCGCTGTGTGCTCCAGGAGGGCCATAGTGTGGCGGGCGGCGCA 457
DB 1456 CATGACCAACATCCAGTCTGTGTCGGGAGGGGCGCCACCGTGTATGCCACAGGACCGCA 1397
QY 458 COTGGGGAGCAGCAGACCAAGCTCCTGCAGCATCTGTGGAGGCTGTGGCGGGTGTGA 517
DB 1396 CGCCAGGTGGAGGAGCGGAGGCTCATGGAGCAGCTGTGAGCAGCTGGGCTTCTGCAC 1337
QY 518 GGAGGTGCTGAAGCTTACGTCAGATCCACACTGGCTCAGTGGCAGTGGGGTGGCTT 577
DB 1336 GGAGGTGGAAGAGGAGCTGATTTGATGCGCTACCGGGGCTCAGTGGCAGCGCGCCGCTA 1277
QY 578 CGTGTGTGATTTCTCCGAGGCGCTTGGCTGAAGGAGCGCTCAAGATGGGCATGCCACGAG 637
DB 1276 CGCATTCACAGCCCTGGATGCCCTGCTGATGGGGGTGTGAAGTGGGATTCGAAGCGG 1217
QY 638 COTGGCCCCACCGATCGCTGCCAGACCGCTGTGGGAGCGGCCAAGATGCTGCTGCACGA 697
DB 1216 COTGGCAGTCCGCTCGGGGCGCCAGCGCCCTCTCGGGGCTGCCAAGATGCTGCTGCACTC 1157
QY 698 GGGCCAAACACCCAGCCAGCTCGCTCAGACCTGTGCACCCCGGGTGGCACCACCATCTA 757
DB 1156 AGAACAGCACCAGCGCAGCTCAAGGACAACTGACGTCTCTGTTGGGGCGCCACCATCA 1097
QY 758 TGGACTCCACGCGCTGAGCAGGGGGGTGGAGCAGCAGCATCATGAGCGCGTGGAGGC 817
DB 1096 TGCCTTGATGTGCTGGAGAGTGGGGGCTTCGCTCCCTGCTCATCAGCGCTGTGGAGGC 1037
QY 818 TGCCACCTCCCGGGCGCAAGGAGCT 841

DB 1036 CTCCTGCATCCGCACACGGGAGCT 1013

RESULT 12

AAF15843

ID AAF15843 standard; cDNA; 1478 BP.

XX AAF15843;

AC AAF15843;

XX 13-MAR-2001 (first entry)

DT Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

DE Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:278.

QY 335 GCTGGGGTCTCTGAGCACCCTGAGGAGCTGCTG-----CCCCAAACACACGG 386
DB 480 GCTGGTGTACCATCAGCTCTGTGAGAAGACTGGGATTCAGCCAGCCCCAAA 539
QY 387 GTGCTGCGGGTCTTCCCAACCTGCGCTGTGTGTCAGAGAGGGCCATAGTGTAGCGC 446
DB 540 GTGATTCGCTGATGACCAACACACCTGTGTGTAGTGCAGGAGGGCTACAGTGTACGCC 599
QY 447 CGGGGCGCGCAGCTGGGAGCAGCAGACCAAGCTCTGTCGACAGCTCTGCTGGAGCCGTGT 506
DB 600 ACGGGCACCACCTGCGCTGTGTGAGGATGGCAGCTCTGTCGAGCAGCTCATGAGCAGCGTG 659
QY 507 GGGCGGTGTGAGAGGTGCTGAAGCCTAGCTGACATCCACACTGCGCTCAGTGGCAGT 566
DB 660 GCTTCTGACTGAGGTGAAGAGACCTCATGATGCCGTACGGGGCTCAGTGGCAGC 719
QY 567 GGGCTGGCCCTGCTGTGTGTCATCTCCGAGGCGCTGGCTGAAGGAGCGCTCAAGATGGGC 626
DB 720 GGGCTGCGCTATGCAATTCATGCTCTGGACGCAATGGCTGATGGTGGGTGAAGATGGT 779
QY 627 ATGCCAGCAGCTGGCCACCGCATCGCTGCGCCAGACCTGCTGGGAGCGCCCAAGATG 686
DB 780 TTGCCACGGCGCTGGCAATCAACTCGGGGGCCAGGCTTTGCTGGGAGCTGCCAAGATG 839
QY 687 CTGCTGACGAGGGCCCAACACCGCAGCTGCGCTCAGACGCTGTGCACCGCGGGTGGC 746
DB 840 CTGCTGACTGAGGAGCAGCTCCATGCGCAGCTTAAGGACATGTCTGCTCCCTGGGGGA 899
QY 747 ACCACCATATGAGCTCCAGCCCTGGAGCAGGGGGCTGCGAGCAGCCACCATGAGC 806
DB 900 GCCACCATCCAGCCCTGCACCTTTCTAGAGAGTGGGGCTTCGCTCTCTGCTCATCAAT 959
QY 807 GCGGTGGAGCTGCCACCTGCGGGGCCCAAGAGCT 841
DB 960 GCAGTTGAGGCTCTGTATCCGAACACGAGAGCT 994

RESULT 14

AAD20894
ID AAD20894 standard; cDNA; 1742 BP.
XX
AC AAD20894;
XX
DT 15-JAN-2002 (first entry)
XX
DE Human delta 1-pyrroline-5-carboxylate reductase homologue (P5CRH) cDNA.
XX
KW Human; delta 1-pyrroline-5-carboxylate reductase; P5CR; cystic fibrosis;
KW osteoporosis; neuronal disorder; gene therapy; akathisia; drug screening;
KW actinic keratosis; Alzheimer's disease; amyotrophic lateral sclerosis;
KW connective tissue disorder; myocardial fibrosis; cell proliferation;
KW arteriosclerosis bursitis; cancer; amnesia; neuroprotective; cytosstatic;
KW antiarteriosclerotic; osteopathic; cardiac; ss.
XX
OS Homo sapiens.
XX
PH Key Location/Qualifiers
FT CDS 192..1154
FT /*tag= a
FT /product= "Human P5CRH"
FT sig_peptide 192..245
FT /*tag= b
FT mat_peptide 246..1151
FT /*tag= c
FT /product= "Mature human P5CRH"
XX
XX US6268192-B1.
XX
XX
XX
PD 31-JUL-2001.
XX
XX 05-MAY-2000; 2000US-0565910.
XX
XX

PR 18-JUN-1998; 98US-0099676.
XX (INCY-) INCYTE GENOMICS INC.
XX Hillman JL, Corley NC, Baughn MR;
XX WPI: 2001-647056/74.
XX P-PSDB; RAE12784.
XX
XX New delta 1-pyrroline-5-carboxylate reductase polypeptides and
XX polynucleotides, useful for diagnosing, treating and preventing
XX neuronal disorders, connective tissue disorders or disorders of cell
XX proliferation
XX
XX Example 5; Fig 1; 32pp; English.
XX
XX The present invention relates to delta 1-pyrroline-5-carboxylate
XX reductase (P5CR) polypeptides and polynucleotides. The P5CR gene is used
XX in gene therapy. The P5CR nucleic acid and amino acid sequences are
XX useful in the diagnosis, treatment and prevention of neuronal disorders
XX (e.g. akathisia, Alzheimer's disease, amnesia, or amyotrophic lateral
XX sclerosis), connective tissue disorders (e.g. cystic fibrosis, myocardial
XX fibrosis, or osteoporosis) and disorders of cell proliferation (e.g.
XX actinic keratosis, arteriosclerosis bursitis, cancers). P5CR may also be
XX used for screening libraries of compounds in various drug screening
XX techniques. The present sequence is a cDNA encoding human delta 1-
XX pyrroline-5-carboxylate reductase homologue (P5CRH).
XX
XX Sequence 1742 BP; 385 A; 490 C; 504 G; 363 T; 0 other;
XX
XX Query Match 15,1%; Score 201.8; DB 22; Length 1742;
XX Best Local Similarity 55,0%; Pred. No. 1.6e-35;
XX Matches 448; Conservative 0; Mismatches 352; Indels 15; Gaps 2;
XX
QY 42 GAGCCGCTCCCGCGCGGTTCGTGGCGCGGCGCATGGCGGGGGCCATCGCG 101
DB 180 GAACCGGGGACCATGAGCGTGGGCTTCATCGGGGCGCGGCTATGCTCTGGCG 239
QY 102 CAGGCGCTCATCAGAGCAGGAAAGTGAAGCTCAGACATCTACTGGCAGTGCAACAA 161
DB 240 CGGGGCTTCAGCGCGCAGGATCTCTGCGCTCACAGATATATAGCAGCTCCCAAG 299
QY 162 GACAGGAACC-----TATGTCACTTTCAAGCTCTGGGTTCGCGGACCAAGCACTCAAC 215
DB 300 ATGAACCTGCCACCGGTTCGCGCTCAGGAAGATGGTGTGAACCTGACACGAGCAAC 359
QY 216 CAGAGGTGTGAGAGCTGCTGCTCATCTTTGCCACCAGACCTCATGTCTGCCA 275
DB 360 AAGGAGCGGTGAAGCAGCAGCAGCTCTGTTCTGGCTGTGAAGCCACATATCATCCCC 419
QY 276 GCTGTCTTCAGAGGTGGCTCTCTGTGTGTCACCACTGAACACATCTTGGTGTCTGGT 335
DB 420 TTATCTCTGATGATGTTGGGCGCGACGTGCAAGCCAGACACATCTGGTCTCTGTGGC 479
QY 336 GCTGGGTGTCTCTGAGCACCTTGAGAGCTGCTG-----CCCCAAACACACGG 386
DB 480 GCTGTGTCAACCATCAGCTCTGTGGAAGAAGCTGTATGGCATTCAGCCAGCCCCAAA 539
QY 387 GTGCTGGGGTCTTGGCCCAACCTGCTGCTGTCAGGAGGAGGGCCATAGTGTGGC 446
DB 540 GTGATTCGCTGCTGACCAACACACCTGTGTGTAGTGCAGGAGGGCTACAGTGTACGCC 599
QY 447 CGGGGCGCGCAGCTGGGAGCAGCAGACCAAGCTCTCTGACAGCTCTGCTGGAGCGCTGT 506
DB 600 ACGGCAACCATGCGCTGGTGGAGATGGGAGCTCTCTGGAGAGCTCATGAGCAGCGTG 659
QY 507 GGGCGGTGTGAGAGGTGCTTGAAGCTTACGTGCACATFCCACACTGGCCCTCAGTGGCAGT 566
DB 660 GGCTTCTGCACTGAGGTGAAGAGACCTCATGATGCCGTGCGGCGCTCAGTGGCAGC 719
QY 567 GCGTGGCCCTTCGTGTGTGTCATCTCCGAGGCGCTTGGCTGAAGAGCGCTCAAGATGGGC 626
DB 720 GGGCTGCGCTATGCAATTCATGCTCTGGACGCAATGGCTGATGGTGGGTGAAGATGGGT 779

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 03:09:38 ; Search time 3560 Seconds
(without alignments)
10938.077 Million cell updates/sec

Title: US-09-806-536A-29
Perfect score: 1338
Sequence: 1 ggtgagcgagctgtctcga.....ggggcagcagcatatggggt 1338

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rtd.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1307	97.7	2331	9	AK023914	AK023914 Homo sapi
C	1291.8	96.5	2615	9	AK001500	AK001500 Homo sapi
3	1200.4	89.7	2400	9	HS0805145	AL833857 Homo sapi
4	1127.8	84.3	1178	9	BC007993	BC007993 Homo sapi
C	656.2	49.0	108980	2	AC105210	AC105210 Homo sapi
C	653.2	48.8	164959	2	AC087823	AC087823 Homo sapi
C	653.2	48.8	174906	2	AC067930	AC067930 Homo sapi
8	599.4	44.8	207636	2	AC019122	AC019122 Homo sapi
9	567.6	42.4	1276	10	BC026536	BC026536 Mus muscu
10	338.4	25.3	165459	2	AC120593	AC120593 Rattus no
11	226.4	16.9	1848	9	BC001504	BC001504 Homo sapi
12	224.8	16.8	1792	6	AX409754	AX409754 Sequence
13	224.8	16.8	1792	9	HUMP5CR	M77836 Human pyro
14	213.8	16.0	1769	9	AF218000	AF218000 Homo sapi
15	205.4	15.4	1757	9	BC022244	BC022244 Homo sapi
16	201.8	15.1	1676	9	BC014868	BC014868 Homo sapi
17	201.8	15.1	1708	9	BC020553	BC020553 Homo sapi
18	191.6	14.3	1590	10	BC006882	BC006882 Mus muscu
19	173.4	13.0	1820	10	BC006727	BC006727 Mus muscu
20	167	12.5	999	9	AF151351	AF151351 Homo sapi
C	151.6	11.3	16911	1	SCE7	AL049819 Streptomy
22	148.8	11.1	3149	9	AK001508	AK001508 Homo sapi
23	147	11.0	145544	2	AC116520	AC116520 Mus muscu
C	143.4	10.7	10490	1	AE013306	AE013306 Methanosa
25	143.2	10.7	162879	2	AC120830	AC120830 Rattus no
C	143.2	10.7	196552	2	AC126537	AC126537 Rattus no
27	141.4	10.6	999	3	AY071558	AY071558 Drosophil
28	138.4	10.3	5139	1	AF305580	AF305580 Methanosa
29	138.4	10.3	10357	1	AE011122	AE011122 Methanosa
C	127.6	9.5	28128	2	AC015301	AC015301 Drosophil
31	127.6	9.5	171612	3	AC009461	AC009461 Drosophil
32	127.6	9.5	175440	3	AC007804	AC007804 Drosophil
33	127.6	9.5	253176	3	AE003718	AE003718 Drosophil
34	126	9.4	10029	1	AE012787	AE012787 Chlorobiu
35	123	9.2	2678	1	TTHPROC	D25413 Thermus the
36	116.6	8.7	1089	3	AF098020	AF098020 Drosophil
37	111.8	8.4	4290	3	AF170829	AF170829 Drosophil
38	110.4	8.3	203050	1	AL646071	AL646071 Ralstonia
C	110.2	8.2	45438	2	AC017841	AC017841 Drosophil
40	110.2	8.2	153048	3	AC007891	AC007891 Drosophil
41	110.2	8.2	204173	3	AC007814	AC007814 Drosophil
42	110.2	8.2	232380	3	AE003724	AE003724 Drosophil
43	107.6	8.0	1017	8	AF466195	AF466195 Phytophth
44	107	8.0	164959	2	AC087823	AC087823 Homo sapi
45	105.2	7.9	1356	3	AF241824	AF241824 Leishmani

ALIGNMENTS

RESULT 1
AK023914
LOCUS AK023914 2331 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ13852 fis, clone THY0100934, weakly similar
to PYRROLINE-5-CARBOXYLATE REDUCTASE (EC 1.5.1.2).
ACCESSION AK023914
VERSION AK023914.1 GI:10435905
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens thyroid gland cDNA to mRNA, clone_lib:THY01
clone:THY0100934.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Isoqai, T., Ota, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
Nishikawa, T., Nagai, K., Sugano, S., Ishibashi, T., Fujimori, K.,
Tanai, H., Kinata, M., Watanabe, M., Hiraoka, S., Ishii, S., Kawai, Y.,
Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagahara, K.,
Masuho, Y., and Kanehori, K.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2615)
AUTHORS Isoqai, T. and Otsuki, T.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Takao Isoqai, Helix Research Institute,
Genomics Laboratory; 1332-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction: 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
1..2615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP2005690"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP2"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
BASE COUNT 492 a 831 c 818 g 474 t
ORIGIN
Query Match 96.5%; Score 1291.8; DB 9; Length 2615;
Best Local Similarity 99.1%; Pred. No. 1.7e-221;
Matches 1299; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 27 AAGATGGCAGCTGCGAGCGCTCCGCGGCGCGTTCGTGGCGCGGGCGGCATG 86
DB 2615 AAGATGGCAGCTGCGAGCGCTCCGCGGCGCGTTCGTGGCGCGGGCGGCATG 2556
QY 87 GCGGGGCGCATCGCAGCGGCTCATCAGCAGGAGAAAGTGGAGCTCAGCACAATCTG 146
DB 2555 GCGGGGCGCATCGCAGCGGCTCATCAGCAGGAGAAAGTGGAGCTCAGCACAATCTG 2496
QY 147 GCCAGTGCCACCAACAGACAGGAACCTATGTCACTTTCAAGCTCTGGGTGGCGGACACG 206
DB 2495 GCCAGTGCCACCAACAGACAGGAACCTATGTCACTTTCAAGCTCTGGGTGGCGGACACG 2436
QY 207 CACTCAACAGGAGGTGTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 266
DB 2435 CACTCAACAGGAGGTGTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2376
QY 267 GTGCTGCCAGCTGCTCGCAGAGTGGCTCCTGTGTGCACACTGAACACATCTTGCTG 326
DB 2375 GTGCTGCCAGCTGCTCGCAGAGTGGCTCCTGTGTGCACACTGAACACATCTTGCTG 2316
QY 327 TCCGTGGCTGCTGGGGTGTCTTGAGCACCTTGGAGGAGCTGCTGCCCGCCAAACACAGG 386
DB 2315 TCCGTGGCTGCTGGGGTGTCTTGAGCACCTTGGAGGAGCTGCTGCCCGCCAAACACAGG 2256
QY 387 GTGCTCGGGTCTTGCCCAACTGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 446
DB 2255 GTGCTCGGGTCTTGCCCAACTGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2196
QY 447 CGGGGCGCGCAGTGGGGAGCAGCAGACCAAGCTCCTGCAGCATCTGCTGGAGCGCTGT 506
DB 2195 CGGGGCGCGCAGTGGGGAGCAGCAGACCAAGCTCCTGCAGCATCTGCTGGAGCGCTGT 2136

QY 507 GGGCGGTGTGAGAGGTGCTTGAAGCCTAGTCGACATCCACACTGGCCTCAGTGGCAGT 566
DB 2135 GGGCGGTGTGAGAGGTGCTTGAAGCCTAGTCGACATCCACACTGGCCTCAGTGGCAGT 2076
QY 567 GGGCGGTGTGAGAGGTGCTTGAAGCCTAGTCGACATCCACACTGGCCTCAGTGGCAGT 626
DB 2075 GGGCGGTGTGAGAGGTGCTTGAAGCCTAGTCGACATCCACACTGGCCTCAGTGGCAGT 2016
QY 627 ATGCCAGCAGCCTGGCCCAACCCATCTGCTGCCAGCCTGCTGGGGAGGCGCCAGATG 686
DB 2015 ATGCCAGCAGCCTGGCCCAACCCATCTGCTGCCAGCCTGCTGGGGAGGCGCCAGATG 1956
QY 687 CTGCTGCACAGAGGCGCCAAACACCCAGCCAGCTCGCTGCAGCAGTGTGCACCCCGGTGGC 746
DB 1955 CTGCTGCACAGAGGCGCCAAACACCCAGCCAGCTCGCTGCAGCAGTGTGCACCCCGGTGGC 1896
QY 747 ACCACCATCTATGAGCTTCCACGCCCTGGAGCAGGGGGCTGGAGCAGCAGCAGCAGT 806
DB 1895 ACCACCATCTATGAGCTTCCACGCCCTGGAGCAGGGGGCTGGAGCAGCAGCAGCAGT 1836
QY 807 GCGGTGAGGCTGCCACCTGCCGGCCAAAGAGCTCAGCAGAAAGTAGGCTGGGCTGTG 866
DB 1835 GCGGTGAGGCTGCCACCTGCCGGCCAAAGAGCTCAGCAGAAAGTAGGCTGGGCTGTG 1776
QY 867 CCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 926
DB 1775 CCATCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1716
QY 927 GCTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 986
DB 1715 GCTCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1656
QY 987 CAGGGGCGAGACTTGGAGGTTCCAGCAGCGGGGAGCCCGAGCAGGCGGAGCAGT 1046
DB 1655 CAGGGGCGAGACTTGGAGGTTCCAGCAGCGGGGAGCCCGAGCAGGCGGAGCAGT 1596
QY 1047 CTCCCTCCCGAGTGAGCAGAAAGCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1106
DB 1595 CTCCCTCCCGAGTGAGCAGAAAGCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1536
QY 1107 CCACCTTGTGTCAACATTTGTTTGTAGGGGCCCCAAGAGATGGGCTTGTGTGTGTGTGT 1166
DB 1535 CCACCTTGTGTCAACATTTGTTTGTAGGGGCCCCAAGAGATGGGCTTGTGTGTGTGTGT 1476
QY 1167 GCATGGTGTGGCAGTGTGGTTGAGGCCATGAACAGAACTTACGGTAACAGCAGCGTGGC 1226
DB 1475 GCATGGTGTGGCAGTGTGGTTGAGGCCATGAACAGAACTTACGGTAACAGCAGCGTGGC 1416
QY 1227 CCAATGCTGCTGTGAGCTGGAGCTTGGCTTGTGGCTTTCAGTGTGGGCTCGTGCAGCTA 1286
DB 1415 CCAATGCTGCTGTGAGCTGGAGCTTGGCTTGTGGCTTTCAGTGTGGGCTCGTGCAGCTA 1356
QY 1287 CAGCAGCGCGGCTGCTCATCTCAGCTCTAGGGGCGCAGCAGCATATGGG 1337
DB 1355 CAGCAGCGCGGCTGCTCATCTCAGCTCTAGGGGCGCAGCAGCATATGGG 1305
RESULT 3
HSM805145 HSM805145 2400 bp mRNA linear PRI 12-JUL-2002
LOCUS Homo sapiens mRNA; cDNA DKF2p/61H0716 (from clone DKF2p/61H0716).
DEFINITION AL833857
ACCESSION AL833857.1 GI:21739340
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2400)
AUTHORS Ansorge, W., Winkler, U., Mewes, H.W., Weil, B. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2002) GSF, Institut fuer Bioinformatik MIPS,

COMMENT
Ingolstaedter Landstr. 1, 85764 Neuherberg, Germany
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp761H0716) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
source
Location/Qualifiers
1..2400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="DKFZp761H0716"
/tissue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2400
/gene="DKFZp761H0716"
1..735
/gene="DKFZp761H0716"
/note="similarity to pyroline 5-carboxylate reductase
isoform (Homo sapiens)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD38716.1"
/db_xref="GI:21739341"
/translation="GKVEAQHILASAPDRNLCFHQALGCRTHNQEVLOSCLLVIF
EKAIVMARHVGVSSETNLLHLEACRCEVEPEAYVDIHTLGSVGAFVCAFS
LAEGAVGMGSPSLAHRIAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGLHAL
EOGGLRAATMSAVEATRAKELSRK"
2381
polyA_site
BASE COUNT 452 a 756 c 740 g 452 t
ORIGIN

Query Match 89.7%; Score 1200.4; DB 9; Length 2400;
Best Local Similarity 99.1%; Pred. No. 3.9e-205;
Matches 1207; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 120 GGAAGAGTGGAGCTCAGACATATCTGGCCAGTCACCAACAGAGAACTATGTAC 179
DB 1 GGAAGAGTGGAGCTCAGACATATCTGGCCAGTCACCAACAGAGAACTATGTAC 60
QY 180 TTTCAGCTCTGGGTTGCCGACCAACAGCTGCTGCCAGTGTCTGCCAGAGTGCCT 239
DB 61 TTTCAGCTCTGGGTTGCCGACCAACAGCTGCTGCCAGTGTCTGCCAGAGTGCCT 120
QY 240 CTCGTCATCTTTGCCACCAAGCTCATGTGCTGCCAGTGTCTGCCAGAGTGCCT 299
DB 121 CTCGTCATCTTTGCCACCAAGCTCATGTGCTGCCAGTGTCTGCCAGAGTGCCT 180
QY 300 GTGGTCACCACTGAACACATCTTGTGTCCTGGTGTCTGTGGGTGTCTGTGAGCACC 359
DB 181 GTGGTCACCACTGAACACATCTTGTGTCCTGGTGTCTGTGGGTGTCTGTGAGCACC 240
QY 360 GAGGAGCTCTGCCGCCCAACACAGGCTGTGTCGGGTGTGTCGGGTGTCTGTGAGCACC 419
DB 241 GAGGAGCTCTGCCGCCCAACACAGGCTGTGTCGGGTGTGTCGGGTGTCTGTGAGCACC 300
QY 420 GTCCAGGAAGGGCCATAGTATGTCGCGCGGGGCCACAGTGGGAGCAGGAGACCAAG 479
DB 301 GTCCAGGAAGGGCCATAGTATGTCGCGCGGGGCCACAGTGGGAGCAGGAGACCAAG 360
QY 480 CTCCTGCAGCATCTCTGGAGCCCTGTGGGGGGTGTGAGAGGTGCTGAGCCCTAGCTC 539
DB 361 CTCCTGCAGCATCTCTGGAGCCCTGTGGGGGGTGTGAGAGGTGCTGAGCCCTAGCTC 420

QY 540 GACATCCACACTGGCCTCAGTGGCAGTGGCTGGCTTCGTGTGTGATCTTCGAGGCC 599
DB 421 GACATCCACACTGGCCTCAGTGGCAGTGGCTGGCTTCGTGTGTGATCTTCGAGGCC 480
QY 600 CTGGCTGAAGGAGCGCTCAAGATGGGATGCCAGCCTGGCCACCGCATCGCTGCC 659
DB 481 CTGGCTGAAGGAGCGCTCAAGATGGGATGCCAGCCTGGCCACCGCATCGCTGCC 540
QY 660 CAGACCCCTGCTGGGAGCGCCCAAGATGCTGTGTCACGAGGGCCCAACACCGCCAGCTG 719
DB 541 CAGACCCCTGCTGGGAGCGCCCAAGATGCTGTGTCACGAGGGCCCAACACCGCCAGCTG 600
QY 720 CGCTCAGACGTGTGCACCCCGGGTGGACACCATCTATGGACTTCCACGCCCTGGAGCAG 779
DB 601 CGCTCAGACGTGTGCACCCCGGGTGGACACCATCTATGGACTTCCACGCCCTGGAGCAG 660
QY 780 GCGGGCTGCCAGCAGCCACCATGAGCGCGTGGAGGCTGCCACTGCCGGGCCAAGGAG 839
DB 661 GCGGGCTGCCAGCAGCCACCATGAGCGCGTGGAGGCTGCCACTGCCGGGCCAAGGAG 720
QY 840 CTCAGCAGAAAGTAGGCTGGGCTCTGGCCTATCCCTTCTGCTGCTGTCGCTTCTCCTC 899
DB 721 CTCAGCAGAAAGTAGGCTGGGCTCTGGCCTATCCCTTCTGCTGCTGTCGCTTCTCCTC 780
QY 900 CCTGTGTCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 959
DB 781 CCTGTGTCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 840
QY 960 CTCCTTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 1019
DB 841 CTCCTTCTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 900
QY 1020 GGGGAGCCCGCAGCAGTGGGACACTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1079
DB 901 GGGGAGCCCGCAGCAGTGGGACACTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 960
QY 1080 GTGGCTCTGCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 1139
DB 961 GTGGCTCTGCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 1020
QY 1140 AAGAGATGGCTCTTGGTCAATTTCCCGCATGTTGGCAGTGGTGGAGCCATCAACA 1199
DB 1021 AAGAGATGGCTCTTGGTCAATTTCCCGCATGTTGGCAGTGGTGGAGCCATCAACA 1080
QY 1200 GAATTTACGTTAACAGGACAGCGCTGGCCCAATGCCTGCTGGAGCTGGAGCTTGCCTTT 1259
DB 1081 GAATTTACGTTAACAGGACAGCGCTGGCCCAATGCCTGCTGGAGCTGGAGCTTGCCTTT 1140
QY 1260 GGCTTTCCAAAGTGGGCTGTCAGCTACAGCCAGCGCGGCTGCTCATCTCAGCTTAGG 1319
DB 1141 GGCTTTCCAAAGTGGGCTGTCAGCTACAGCCAGCGCGGCTGCTCATCTCAGCTTAGG 1260
QY 1320 GGGCAGCAGCATATGGG 1337
DB 1201 GGGCAGCAGCATATGGG 1218

RESULT 4
BC007993
LOCUS
DEFINITION Homo sapiens, hypothetical protein FLJ13852, clone MGC:14985
IMAGE:3533609, mRNA, complete cds.
ACCESSION BC007993
VERSION BC007993.1 GI:14124939
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1178)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (15-MAY-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mqc.nci.nih.gov>

Contact: MGC help desk

Email: cqapbs-r@mail.nih.gov

Tissue Procurement: DCTD/DTP

CDNA Library Preparation: Rubin Laboratory

CDNA Library Preparation: KUBIN LABORATORY
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

CDNA LIBRARY ARRAYED BY: THE I.M.A.G.E. CENTRE
DNA SEQUENCING BY: GENOME SEQUENCE CENTRE

BC Cancer Agency

BC cancer agency, Vancouver, BC, Canada
 info@bcsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Len Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saesedi, Jacqueline
 Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasia van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL plate: 23 Row: 1 Column: 8

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA q1: 10435995.

FEATURES

source

```
location/Vascularities
1. .ll78
/dbgenism="Homo sapiens"
/dbxref="LocusID:65263"
/db_xref="taxon:9606"
/clonem="MGC:14985 IMAGE:3533609"
/tissue_type="Lung, small cell carcinoma"
/clonelib="NIH_MGC_7"
/lab_host="DH10B-R"
/note="Vector: pOT87"
47. .871
```

CDS

BASE COUNT

ORIGIN

Query Match	84.3%;	Score 1127.8;	DB 9;	Length 1178;
Best Local Similarity	99.8%;	Pred. No. 4.4e-192;		
Matches 1129; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	15	GTCCGAGCGCAACAGATGGCAGCTGCCGAGCGCTCTCCGGCGCCGCTGGGCTTCGTGGGC	74
Db	32	GTCCGAGCGCAACAGATGGCAGCTGCCGAGCGCTCTCCGGCGCCGCTGGGCTTCGTGGGC	91
Qy	75	CGCGGCCGCATGGCGGGGGCCATCGCGCAGGGCCCTCATCAGAGCAGAAAGTGGAAAGCT	134
Db	92	CGGGCCGCATGGCGGGGGCCATCGCGCAGGGCCCTCATCAGAGCAGAAAGTGGAAAGCT	151
Qy	135	CAGCACATACTGGCCATGTGCACCAACAGACAGAAACCTATGTCTACTTTCAAGCTCTGGGT	194
Db	152	CAGCACATACTGGCCAGTGCACCAACAGACAGAAACCTATGTCTACTTTCAAGCTCTGGGT	211
Qy	195	TGCCGGACACGCACTCCAACAGGAGGTGCTGCAGAGCTGCCCTGCTCATCTTTGGCC	254
Db	212	TGCCGGACACGCACTCCAACAGGAGGTGCTGCAGAGCTGCCCTGCTCATCTTTGGCC	271
Qy	255	ACCAAGCCTCATGTGCTGCCAGCTGTCTCTGGCAGAGGTGGCTCTCTGGTCAACCACTGAA	314
Db	272	ACCAAGCCTCATGTGCTGCCAGCTGTCTCTGGCAGAGGTGGCTCTCTGGTCAACCACTGAA	331

REFERENCE
AUTHORS

2 (bases 1 to 108980)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgaltier, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (26-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 108980)

TITLE
JOURNAL
COMMENT

Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 5, 2002 this sequence version replaced gi:21307003.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L23062
 Center clone name: 118_P12

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 14262: contig of 14262 bp in length
 * 14263 14362: gap of 100 bp
 * 14363 16379: contig of 2017 bp in length

* 16380	16479:	gap of	100 bp
* 16480	18932:	contig of 2453 bp in length	
* 18933	19032:	gap of	100 bp
* 19033	25980:	contig of 6948 bp in length	
* 25981	26080:	gap of	100 bp
* 26081	47787:	contig of 21707 bp in length	
* 47788	47887:	gap of	100 bp
* 47888	82335:	contig of 34448 bp in length	
* 82336	82435:	gap of	100 bp
* 82436	97616:	contig of 15181 bp in length	
* 97617	97716:	gap of	100 bp
* 97717	100356:	contig of 2640 bp in length	
* 100357	100456:	gap of	100 bp
* 100457	105947:	contig of 5491 bp in length	
* 105948	106047:	gap of	100 bp
* 106048	108980:	contig of 2933 bp in length.	

FEATURES
 source
 1. 108980
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8"
 /clone="RP11-118P12"
 /clone_lib="RPC1-11 Human Male BAC"
 BASE COUNT 24096 a 30961 c 30570 g 22231 t 1122 others
 ORIGIN

Query Match	49.0%	Score 656.2;	DB 2;	Length 108980;
Best Local Similarity	93.9%	Pred. No. 4.1e-108;		
Matches 682;	Conservative 0;	Mismatches 44;	Indels 0;	Gaps 0;

QY 612 GCCCTCAAGATGGGCATGCCAGCAGCGCTGGCCACCGCATCGCTGCCAGACCCCTGCTG 671
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 56072 GCCGCACAGTGGGGACCGCACCACCTCCCGCTCACAGGCCCTCAACGCTCTCCTCAG 56013
 QY 672 GGGACGCCAAGATGCTGCTGCAGGAGGGCCACACCCAGCCAGCTGCCTCAGACGTTG 731
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 56012 GGGACGCCAAGATGCTGCTGCAGGAGGGCCACACCCAGCCAGCTGCCTCAGACGTTG 55953
 QY 732 TGACACCCCGGTGGCCACCACTATATGAGCTCCAGCCCTGGAGAGGGCGGCTGCCGA 791
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 55952 TGCACCCCGGTGGCCACCACTATATGAGCTCCAGCCCTGGAGAGGGCGGCTGCCGA 55893
 QY 792 GCAGCCACCATGAGCGCGGTGGAGGCTGCACCTGCGGGCCAAAGAGCTCAGCAAAAG 851
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 55892 GCAGCCACCATGAGCGCGGTGGAGGCTGCACCTGCGGGCCAAAGAGCTCAGCAAAAG 55833
 QY 852 TAGGCTGGGCTTGGCCATCTTTCTGCTCTGTGCCCCCTGCTCTCCCTGTGTCCTTT 911
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 55832 TAGGCTGGGCTTGGCCATCTTTCTGCTCTGTGCCCCCTGCTCTCCCTGTGTCCTTT 55773
 QY 912 CCCTGAGGACTGGGCTCCCTCCCTCCTGCATGAGGCTCTCTACTGCTCTCTCTCCCT 971
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 55772 CCCTGAGGACTGGGCTCCCTCCCTCCTGCATGAGGCTCTCTACTGCTCTCTCTCCCT 55713
 QY 972 TTGCACAGGAAATGAGGGGGCAGGACTTGGAGGTTCCAGAGCGGGGGAGCCCCGA 1031
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 55712 TTGCACAGGAAATGAGGGGGCAGGACTTGGAGGTTCCAGAGCGGGGGAGCCCCGA 55653
 QY 1032 CCAGTGGGGACACTCTCTCCCTCCCGCAGTGAGCAGAGGCACCGTGGTGGCTCTGCC 1091
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 55652 CCAGTGGGGACACTCTCTCCCTCCCGCAGTGAGCAGAGGCACCGTGGTGGCTCTGCC 55593
 QY 1092 CTTGCTGCAGTGAGCCACCTTCTGCTCAACATTTGTTCTGAGGGGCCCAAGAGATGGCT 1151
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 55592 CTTGCTGCAGTGAGCCACCTTCTGCTCAACATTTGTTCTGAGGGGCCCAAGAGATGGCT 55533
 QY 1152 CTTGCTGATTGGCCCGCATGGTTGGGCGAGTTGGTTGAGGCCATGAACAGAACTTACGTA 1211
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 55532 CTTGGTCAATTTGCCCGCATGGTTGGGCGAGTTGGTTGAGGCCATGAACAGAACTTACGTA 55473
 QY 1212 ACAGGCGCGCTGGCCCAATGCTGCTGGAGCTGGAGCTTGCCTTTGCCCTTCCAACT 1271
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

```
Db      55472 ACAGGCACGGCTGCGCCCAATGCTGTCTGGAGCTGGAGCTGCTTTGCTTTCCAGGT 55413
QY      1272 GGGCTCTGTCAGCTACAGCCAGCGCGCTGCCCTCATCTCAGCTCTAGGGGGGACGACCAT 1331
Db      55412 GGCCTCCGTGACGTACAGCCAGCGCGCTGCCCTCATCTCAGCTCTAGGGGGGACGACCCA 55353
QY      1332 ATGGGG 1337
Db      55352 TATGGG 55347

RESULT 6
AC087823/c
LOCUS   AC087823
DEFINITION Homo sapiens chromosome 8 clone RP11-545C16 map 8, *** SEQUENCING
IN PROGRESS ***, 70 unordered pieces.
ACCESSION AC087823.2 GI:13357313
VERSION   1
KEYWORDS  HTG; HTGS_PHASE1.
SOURCE   Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS   1 (bases 1 to 164959)
TITLE     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL   Homo sapiens chromosome 8, clone RP11-545C16
REFERENCE
AUTHORS   2 (bases 1 to 164959)
TITLE     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galdogan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehocsky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pleyre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Trauers,M., Travis,N., Triggilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-JAN-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2001 this sequence version replaced gi:12584321.
All repeats were identified using RepeatMasker:
Smit, A.F.P. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11736
Center clone name: 545_C_16
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```


Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Melrhim, J., Meneses, L., Mihova, R., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Ollivar, T. M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pillara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stege-Thomas, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigliio, J., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A., and Zody, M.

TITLE
JOURNAL

Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 174906)

REFERENCE AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Bogustavsky, L., Boushgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, V., Colangelo, M., Collins, S., Collumore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczyk, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A., Zody, M.

TITLE
JOURNAL

Submitted (21-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
(bases 1 to 174906)

REFERENCE
AUTHORS

4 (bases 1 to 174906)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allien, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boquslavskiy, L., Boukhgaltier, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Keller, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Minova, T., Miengwa, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodor, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Willson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Submitted (24-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 24, 2002 this sequence version replaced gi:21490265.

COMMENT

On Jul 24, 2002 this sequence version replaced g1:21490265.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

Smith, A.F.A. & Green, P. (1990-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome
Center code: WIBR

Web site: <http://www-seq.wi.mi>
Contact: sequence_submissions@
----- Project Information

Center Project name: L10100	
Center clone name: 661_A12	

FEATURES	Location/Qualifiers
Source	1. 174906
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="8"
	/map="8"
	/clone="RP11-661A12"
	/clone_lib="RP11-661A12 Human Male BAC"
repeat_region	complement(351..430)
	/rpt_family="MIR"
repeat_region	977..1108
	/rpt_family="LIME4a"
repeat_region	complement(1149..1449)
	/rpt_family="AluSg"
repeat_region	3675..3781
	/rpt_family="MIR"
repeat_region	complement(16671..16974)
	/rpt_family="AluSg"
repeat_region	20174..20412
	/rpt_family="L2"
repeat_region	20487..20585
	/rpt_family="L2"
unsure	complement(20694..20702)
	/note="<30 qual SINGL region"
repeat_region	23655..23772
	/rpt_family="LIME2"
repeat_region	24264..24383
	/rpt_family="LIME2"
repeat_region	complement(24446..24779)
	/rpt_family="MLT1B"
unsure	26300
	/note="probably A; possibly G"
repeat_region	26881..26825
	/rpt_family="LIME2"
repeat_region	28798..29091
	/rpt_family="AluX"
repeat_region	32096..32118
	/rpt_family="AT-rich"
repeat_region	32990..33157
	/rpt_family="LIME2"
repeat_region	33598..33749
	/rpt_family="LIME2"
repeat_region	complement(33750..34064)
	/rpt_family="AluJo"
repeat_region	34065..34129
	/rpt_family="LIME2"
repeat_region	34132..34706
	/rpt_family="LIME4a"
repeat_region	34991..35084
	/rpt_family="LIME3"
repeat_region	35191..35325
	/rpt_family="LIME4"
repeat_region	36901..37069
	/rpt_family="(CGGG)n"
repeat_region	37077..37162
	/rpt_family="(CAGG)n"
repeat_region	complement(38672..38844)
	/rpt_family="MSTRD"
unsure	complement(38859..38884)
	/note="<30 qual SINGL region"
repeat_region	complement(38880..38983)
	/rpt_family="MSTRC"
unsure	complement(38896..38903)
	/note="<30 qual SINGL region"
unsure	complement(38898..38902)
	/note="probably ACTGA"
repeat_region	complement(38984..39281)
	/rpt_family="AluX"
repeat_region	complement(39282..39334)
	/rpt_family="MSTRC"

```
repeat_region complement(39366..39668)
repeat_region /rpt_family="AluJb"
repeat_region /rpt_family="MLT1A2"
repeat_region 41752..41933
repeat_region /rpt_family="(CCA)n"
repeat_region 42478..42510
unsure /rpt_family="G-rich"
unsure complement(43664..43676)
unsure /note="<30 qual SNGL region"
unsure complement(43666..43674)
unsure /note="probably ACTGGTGGC"
unsure 46481..46483
repeat_region /note="probably CCC"
repeat_region 54691..54749
repeat_region /rpt_family="(CA)n"
repeat_region 56353..56388
repeat_region /rpt_family="MER5B"
repeat_region 60486..60780
repeat_region /rpt_family="AluSx"
repeat_region 60802..61091
repeat_region /rpt_family="AluJb"
repeat_region complement(69122..69423)
repeat_region /rpt_family="AluJb"
repeat_region 71273..71458
repeat_region /rpt_family="L2"
repeat_region 71493..71733
repeat_region /rpt_family="L2"
repeat_region complement(71849..72140)
repeat_region /rpt_family="AluSx"
repeat_region 72167..72461

Query Match 48.8%; Score 653.2; DB 9; Length 174906;
Best Local Similarity 98.1%; Pred. No. 1.3e-107;
Matches 661; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 664 CCTCTGGGGACGCCAAGATGCTGTCAGAGGGCCACACCCAGCCAGCTGGCGT 723
DB 144055 TCCTCTAGGGGACGCCAAGATGCTGTCAGAGGGCCACACCCAGCCAGCTGGCGT 143996

QY 724 CAGAGGTGTGACCCCGGGTGGCCACCATCTATGGACTCCAGCCCTGGAGCAGGGCG 783
DB 143995 CAGAGGTGTGACCCCGGGTGGCCACCATCTATGGACTCCAGCCCTGGAGCAGGGCG 143936

QY 784 GCCTGGAGACCCACCATGAGCGCGTGGAGGTGCCACCTGCCGGGCCAAGAGCTCA 843
DB 143935 GCCTGGAGACCCACCATGAGCGCGTGGAGGTGCCACCTGCCGGGCCAAGAGCTCA 143876

QY 844 GCAGAAAGTAGGCTGGGCTCTGGCCATCTTTCTGCTCTGTGCCCTGCTCTCCCTG 903
DB 143875 GCAGAAAGTAGGCTGGGCTCTGGCCATCTTTCTGCTCTGTGCCCTGCTCTCCCTG 143816

QY 904 TGTCCCTTCCCTTAGGACTCGGCTCCCTCCCTCCCTGCATGAGGGTCTCTACTGTCC 963
DB 143815 TGTCCCTTCCCTTAGGACTCGGCTCCCTCCCTCCCTGCATGAGGGTCTCTACTGTCC 143756

QY 964 TTCTCCCTTCGACAGGGAATGACAGGGGCGAGACTTGGAGGTTCCAGCAGCGGGGG 1023
DB 143755 TTCTCCCTTCGACAGGGAATGACAGGGGCGAGACTTGGAGGTTCCAGCAGCGGGGG 143696

QY 1024 AGCCCCGACCACTAGTGGGGACACTCTCTCCCTCCCTAGTGAGCAGAGGACCGTGGTGG 1083
DB 143695 AGCCCCGACCACTAGTGGGGACACTCTCTCCCTCCCTAGTGAGCAGAGGACCGTGGTGG 143636

QY 1084 CTCTGCCCTTCTGTCAGTGAGCCACCTTGTCTCAACATTTGTTCTGAGGGGCCCAAGA 1143
DB 143635 CTCTGCCCTTCTGTCAGTGAGCCACCTTGTCTCAACATTTGTTCTGAGGGGCCCAAGA 143576

QY 1144 GATGGCGTCTGGTTCATTTGGCCCGCATGTTGGGCGAGTTGTTGAGGCCATGACAGAAC 1203
DB 143575 GATGGCGTCTGGTTCATTTGGCCCGCATGTTGGGCGAGTTGTTGAGGCCATGACAGAAC 143516

QY 1204 TTACGGTAACAGGCACGCTGGCCCCAATGCTGCTGGAGCTGGAGCTTGCCTTTGGCT 1263
```

```
Db 143515 TTACGGTAACAGGCACGCTGGCCCCAATGCTGCTGGAGCTGGAGCTTGCCTTTGGCT 143456
QY 1264 TTCCAAGTGGCTCGTCGACGTACAGCCAGCGCGCTGCTCATCTCAGCTCTAGGGGCG 1323
Db 143455 TTCCAGTGGCTCCGTGACGTACAGCCAGCGCGCTGCTCATCTCAGCTCTAGGGGCG 143396
QY 1324 ACGAGCATATGGGG 1337
Db 143395 ACGAGCATATGGG 143382

RESULT 8
AC019122 207636 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 8 clone RP11-545C16, WORKING DRAFT
DEFINITION SEQUENCE, 12 unordered pieces.
AC019122 AC019122.3 GI:8318560
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 207636)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 207636)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 7, 2000 this sequence version replaced gi:7109639.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: L NH0545C16
----- Summary Statistics -----
Sequencing vector: M13; 88%
Sequencing vector: plasmid; 12%
Chemistry: Dye-primer ET; 88% of reads
Chemistry: Dye-terminator Big Dye; 12% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 198432 bases at least Q40
Consensus quality: 201424 bases at least Q30
Consensus quality: 203070 bases at least Q20
Insert size: 182000; agarose-fp
Quality coverage: 6.36 in Q20 bases; agarose-fp
Quality coverage: 5.66 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1653: contig of 1653 bp in length
* 1654 1753: gap of unknown length
* 1754 3687: contig of 1934 bp in length
* 3688 3787: gap of unknown length
* 3788 11012: contig of 7225 bp in length
* 11013 11112: gap of unknown length
* 11113 16540: contig of 5428 bp in length
* 16541 16541: gap of unknown length
* 23739: contig of 7099 bp in length
```

* 23740 23839: gap of unknown length
* 23840 contig of 11200 bp in length
* 35039 contig of unknown length
* 35040 35139: gap of unknown length
* 35140 54331: contig of 19092 bp in length
* 54232 54331: gap of unknown length
* 54332 78381: contig of 24050 bp in length
* 78382 78481: gap of unknown length
* 78482 106509: contig of 28028 bp in length
* 106510 106609: gap of unknown length
* 106610 130451: contig of 23842 bp in length
* 130452 130551: gap of unknown length
* 130552 157163: contig of 26612 bp in length
* 157164 157263: gap of unknown length
* 157264 207636: contig of 50373 bp in length.

FEATURES

source

1. 207636
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="RP11-545C16"

misc_feature 1. .1653
/note="assembly_name:Contig14"

misc_feature 1754. .3687
/note="assembly_name:Contig15"

misc_feature 3788. .11012
/note="assembly_name:Contig16"

misc_feature 11113. .16540
/note="assembly_name:Contig17"

misc_feature 16641. .23739
/note="assembly_name:Contig18"

misc_feature 23840. .35039
/note="assembly_name:Contig19"

misc_feature 35140. .54231
/note="assembly_name:Contig20"

misc_feature 54332. .78381
/note="assembly_name:Contig21"

misc_feature 78482. .106509
/note="assembly_name:Contig22"

misc_feature 106610. .130451
/note="assembly_name:Contig23"

misc_feature 130552. .157163
/note="assembly_name:Contig24"

misc_feature 157264. .207636
/note="assembly_name:Contig25"

vector_side:left
clone_end:r7

BASE COUNT 44315 a 59454 c 59618 g 43137 t 1112 others
ORIGIN

Query Match 44.8%; Score 599.4; DB 2; Length 207636;
Best Local Similarity 96.9%; Pred. No. 5.2e-98;
Matches 653; Conservative 0; Mismatches 17; Indels 4; Gaps 4;

QY 664 CCTGCTGGGAGCGCCAGATGCTGCTGACAGAGGCCAACCCAGCCAGTCGCGT 723

DB 48388 CTCCTCAGGGGAGCGCCAGATGCTGCTGACAGAGGCCAACCCAGCCAGTCGCGT 48447

QY 724 CAGACGTGTGCACCCCGGTGTCACCACTATGAGTCTCCACCCCTGGAGCAGGGCG 783

DB 48448 CAGACGTGTGCACCCCGGTGTCACCACTATGAGTCTCCACCCCTGGAGCAGGGCGT 48507

QY 784 GGCTGGAGCAGCCACCATGAGCGCGTGGAGGCTGCCACCTGCCGGGCCAAGGAGCTCA 843

DB 48508 GNCCTGGAGCAG-CACCATGAGCGCGTGGAGG-TGCCACCTGCCGGGCCAAGGAGCTCA 48565

QY 844 GCAGAAAGTAGGCTGGGCTCTGGCCATCCTTTCTGCTGCTGTGCCCTGCTCCCTG 903

DB 48566 GCAGAAAGTAGGCTGGGCTCTGG-CATCCTTTCTGCTGCTGTGCCCTGCTCCCTG 48624

QY 904 TGTCCCTTCCCTTGAGGACTGGGCTCCCTCCCTCTCTGATGAGGCTCTCTACTGCTCC 963

DB 48625 TGTCCCTTCCCTTGAGGACTGGGCTCCCTCCCTCTCTGATGAGGCTCTCTACTGCTCC 48684

QY 964 TTCTCCCTTGCACAGGGAATGACGGGGCAGACTTGGAGGTTCACAGCAGCGGGG 1023
DB 48685 TTCTCCCTTGCACAGGGAATGACGGGGCAGACTTGGAGGTTCACAGCAGCGGGG 48744
QY 1024 AGCCCCGACAGTGGGACACTCTCCCTCCCACTGAGCAGAGGACCGTGGTGTGG 1083
DB 48745 AG-CCCGACCACTAGTGGGACACTCTCCCTCCCACTGAGCAGAGGACCGTGGTGTGG 48803
QY 1084 CTCTGCCCTTGTCTGACAGTGGGACACTCTCTCTCAACATTGGTCTGAGGGGCCCAAGA 1143
DB 48804 CTCTGCCCTTGTCTGACAGTGGGACACTCTCTCTCAACATTGGTCTGAGGGGCCCAAGA 48863
QY 1144 GATGGCGTCTTGGTCACTTTGCCGCATGTTGGGACAGTGGTGTGAGGCCATGAACAAAC 1203
DB 48864 GATGGCGTCTTGGTCACTTTGCCGCATGTTGGGACAGTGGTGTGAGGCCATGAACAAAC 48923
QY 1204 TTACGGTAACAGGACAGCGCTGGCCCAATGCTGCTGGAGCTGGAGCTTGCCCTTTGGCT 1263
DB 48924 TTACGGTAACAGGACAGCGCTGGCCCAATGCTGCTGGAGCTGGAGCTTGCCCTTTGGCT 48983
QY 1264 TTCCAAGTGGGCTCGTGCAGCTACAGCCAGCGCGCTGCCCTCATCTCAGCTCTAGGGGGC 1323
DB 48984 TTCCAGGTGGCTCGTGCAGCTACAGCCAGCGCGCTGCCCTCATCTCAGCTCTAGGGGGC 49043
QY 1324 ACGAGCATATGGG 1337
DB 49044 ACGAGCATATGGG 49057

RESULT 9

BC026536

LOCUS

DEFINITION

IMAGE:4987954, mRNA, complete cds.

ACCESSION

BC026536

VERSION

BC026536.1

KEYWORDS

MGC.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 1276)

AUTHORS

Strausberg, R.

TITLE

Direct Submission

JOURNAL

Submitted (02-APR-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

REMARK

COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcdepaxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAP Plate: 58 Row: d Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13384799.
Location/Qualifiers
1. 1276
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="MGC:35759 IMAGE:4987954"
/tissue_type="Colon, normal, 5 month old male mouse."

FEATURES

source

```
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
16. 840
/codon_start=1
/product="RIKEN cDNA 1110058B13 gene"
/protein_id="AAH26536.1"
/db_xref="GI:20071634"
/translation="MAATMSPRRVFVAGRMARAEATARGLIQAGKVEAKOVLASAPT
DNILCFRALGCTGTHSNHEVLQNCPLVIFATKPOVLPTVLAEPVITTEHIIIVSA
AGISLSTWEGLLPNTNRVLRVSPNLCVQBGAMAMKMGHAGNDRELLQNLLEACG
ACIEVPSYVDHITGLSGVAFCTTSEALAGAIMKMPISGLAHRAAATLGLTAK
MLQQRKHPAQLRTDVLTPAGTTHGLHALERGGFRATMSAVEAATCRKELSK"
BASE COUNT 288 a 357 c 360 g 271 t
ORIGIN
Query Match 42.4%; Score 567.6; DB 10; Length 1276;
Best Local Similarity 79.8%; Pred. No. 7.2e-92;
Matches 669; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
QY 24 AACAAAGATGGCAGCTCGCGAGCGCTCCGCGGCGCTGGCTTCGTGGCGCGGCGGC 83
D 1111111111111111111111111111111111111111111111111111111
Db 10 ACAGGATGGCGGCACAACTGTCTGAACCTCGGCTGTGGCTTCGTGGTGCAGCGCC 69
QY 84 ATGGCGGGGCCATCGCGAGGCGCTCATCATGAGCAGGAGAAAGTGGAGCTCAGCACA 143
D 1111111111111111111111111111111111111111111111111111111
Db 70 ATGGCGGGAGCCATTGGCCGAGGCGCTCATCCAAAGCAGGCAAGTAGAAGCTTAAAC 129
QY 144 CTGGCCAGTGCACCAACAGAGAACTATGTCACTTCAAGCTCTGGTTCGGGACC 203
D 1111111111111111111111111111111111111111111111111111111
Db 130 CTGGCCAGTGCACCAACAGGACAACTCTGCCACTTCAGGCTCTGGTTCGGAGCT 189
QY 204 AGCACTCAACAGGAGGTGCTGCAGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCT 263
D 1111111111111111111111111111111111111111111111111111111
Db 190 ACTCACTCAACAGTGGTGTGCAAACTGCCACTTGTCTATCTTTGCCACCAACCC 249
QY 264 CATGTGCTGCCAGCTCTCTGCAGAGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCT 323
D 1111111111111111111111111111111111111111111111111111111
Db 250 CAAGTCTGCCAACTGCTCTGCGGAAGTGGCCCGCCATAGTACCACTGAGCAGCATC 309
QY 324 GTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 383
D 1111111111111111111111111111111111111111111111111111111
Db 310 GTATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 369
QY 384 CGGTGCTGCGGCTCTGCCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 443
D 1111111111111111111111111111111111111111111111111111111
Db 370 CGAGTATTCGAGTCTCTCCCAATCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 429
QY 444 GCGCGGGCGCGCAGTGGGAGCAGCAGACCAAGCTCTCTGCAAGCATCTGCTGAGGCC 503
D 1111111111111111111111111111111111111111111111111111111
Db 430 GCGCGGGCGCCAGTCTGGAACATGACGAGCTCTCTACAGAACTTCTGCTGGAAGCC 489
QY 504 TGTGCGGCTGTGAGGAGTGCCTGAAGCTACGTGCTGACATCCACACTGCGCTCAGTGG 563
D 1111111111111111111111111111111111111111111111111111111
Db 490 TGTGGCAGTGCATAGAGTTCGCGAGTCTCTAGTATATCCACACCGGCTCAGTGGC 549
QY 564 AGTGGCTGCGCTTCTGTGTGCTATCTCCAGGCGCTGCTGAGAGCGCGTCAAGATG 623
D 1111111111111111111111111111111111111111111111111111111
Db 550 AGTGTGTGGCTTTTGTGTGTACATTTTCAGAGGCGCTGCTGAGGTGCCATCAAAAT 609
QY 624 GGCATGCCAGCAGCTGGCCACCGCATCTGCTGCCAGACCTGCTGGGAGCGCCCAAG 683
D 1111111111111111111111111111111111111111111111111111111
Db 610 GGCATGCCAGTGGCTGGCCACCGCATCTGCTGCTAGACCTTCTGCTGGGAGCAGCC 669
QY 684 ATGCTGCTGACAGGGGCCAACCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 743
D 1111111111111111111111111111111111111111111111111111111
Db 670 ATGCTGACAGGAGGAGGAGCAGCCAGCCAGCTTCCGACAGATGTGCTCACACCACT 729
QY 744 GGCACCCACCTATGAGTCTCCAGCCCTGAGCAGGCGGCTGCGCAGCAGCCACCATG 803
D 1111111111111111111111111111111111111111111111111111111
Db 730 GGAACCCACCTATGAGTCTCCAGCCCTGAGCAGGCGGCTTTTCAGCGGCTTACCATG 789
```

```
QY 804 AGCGCGGTGAGGCTGCCACTCGCGGCGCAAGAGCTCAGCAGAAAGTAGGCTGGC 861
D 1111111111111111111111111111111111111111111111111111111
Db 790 AGTGGGTGGAACGACGCTACTCTCGCGGCTAAGGAGCTCAGCAGAAAGTAGGCGC 847

RESULT 10
AC120593 165459 bp DNA linear HTG 24-AUG-2002
LOCUS Rattus norvegicus clone CH230-300L19, *** SEQUENCING IN PROGRESS
DEFINITION *** 58 unordered pieces.
AC120593
VERSION AC120593.3 GI:22164892
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 165459)
AUTHORS Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angulano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyly,M., Cree,A., D'Souza,L.,
Devila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Iolebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuewa,L., Louseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,P., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., Mcleod,M., McNeill,T., Meenen,E., Milosavljevic,A.,
Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G.,
Olariunpagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A.,
Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
Sisson,I., Sitter,C.D., Smajk,D., Sneed,A., Sodergren,E.,
Song,X.-Z., Sorrelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
Syatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlarczyk,R.,
Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
Direct Submission
TITLE
JOURNAL
REFERENCE 2 (bases 1 to 165459)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Human Genome Sequencing Center, Department
```



```
QY 150 AGTGACCAACACAGACAGCAAGCAACATATGTCACATTTCAAGCTCTGGGTTCGGGACCAACGACG 209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138616 AGTGC-----ACAAGAACCTCTGCCACTTAAAGGCTCTGGTTG-----TCAC 138658

QY 210 TCACACAGGAGGTGCTGCAGAGCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138659 TCACACAGGAGGTGCTGCMAAATGCGCACTTGTCTATCTTTGCCACCAAAACCCCAAGTC 138718

QY 270 CTGCCAGCTGCTCTGCCAGAGGTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138719 CTGCCAGCGTCTCTGCCAGAGGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 138778

QY 330 GTGGCTGCTGGGTGCTCTCTGAGCACCCTGGAGGAGCTGCTGCCCCCA-----AACACACG 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138779 GTGGATGCTGGGATCTCTCTGAGCTCTATGGAGGAGCTATTTCCGCCACCCCAACACACA 138838

QY 386 GTGCTGCGGGTCTTGGCCCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 445
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138839 AGTATTTGGAGTCTCTCTCAATCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 138898

QY 446 GCGGGGCCCCACGCTGGGAGCAGCAGACCAAGCTCTGCGAGCATCTGCTGGAGGCGCTG 505
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138899 ACGGGCCACCATGTTGGGAATGATGACGCTAGGCTCTATAGACTTGTATGGAACCTG 138958

QY 506 TGGGGGGTGTGAGGAGGTGCTGGAAGCCTACGTCGACATCCACACTGCGCTCAGTGGCGAG 565
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138959 TGGGCAATTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 139014

QY 566 TGGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139015 TGGTGTGGCCCTCA-----CTGAAGTGGCGTCA--ATGGG 139048

QY 626 CATGCCACGACGCTGGCCACCGCATCGCTGCCAGACCTGCTGGGAGCGGCCAAGAT 685
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139049 CATGCCAGTGGCTGGCCACCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 139108

QY 686 GCTGCTGACGAGGCGCCACACCCAGCCAGCTGCGCTGACAGCTGTCACCCCGGGTGG 745
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139109 GCTGTCAGCAGGAAGGAAGCAGCCAGCCAGCTGCGGACAGATGT-----GCTGG 139158

QY 746 CACCACCATCTATGACTTCCAGCCCTGGAGCAGGCGGGCTGCGAGCAGCCACCATGAG 805
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139159 ATACACCATCCACCGCTTCACACCTAGATCAAGCGCGTTTTCAGCAGCTACCTTGAG 139218

QY 806 CCGCTGTGAGGCTGCCACTGCGGGGCCAAGAGGCTCAGCAGAAAGTAGGTGGGC 861
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139219 TGCTGTGGAAGCAGCTACCTGCTGGCAAGTAGCTCAGCAAGAGTCAGGCAGGC 139274

RESULT 11
BC001504
LOCUS
DEFINITION
Homo sapiens, 1848 bp mRNA linear PRI 29-OCT-2001
IMAGE:3505512, mRNA, complete cds.
ACCESSION
BC001504
VERSION
BC001504.1 GI:16306657
KEYWORDS
MGC.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1848)
Strausberg,R.
Direct Submission
Submitted (21-DEC-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
AUTHORS
Contact: MGC help desk
JOURNAL
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 1769) Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P., Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y., Yu,J. and Han,L.H. Novel Human cDNA clones with function of inhibiting cancer cell growth Unpublished 2 (bases 1 to 1769) Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P., Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y., Yu,J. and Han,L.H. Direct Submission Submitted (21-DEC-1999) National Laboratory For Oncogenes & Related Genes, Shanghai Cancer Institute, 25 Ln 2200, Xie-Tu Road, Shanghai 200032, People's Republic of China Location/Qualifiers																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Db 758 GGAGGTGGAAGAGGACCTGATTGATGCCGCTCAGCGGGCTCAGTGGCAGCGGCCCGCCTA 817
QY 578 CGTGTGTGATCTCCGAGGCCCTGGCTGAAGAGCCGTCAGATGGGCATGCCAGCAG 637
Db 818 CGATTTCAGAGCCCTGGATGGCTGGCTGATGGGGTGTGAAGATGGGACTTCCAAGGCG 877
QY 638 CTGGGCCACCGCATCGCTGCCAGACCCCTGCTGGGGAC-GGCCAAGATGCTGCTGCAGC 696
Db 878 CTGGCAGTCCGCTCGGGGCCAGGCCCTCTCGGGGCTTCCCAAGATGCTGCTGCAC 937
QY 697 AGGCCCAACACCGAGCCCGAGCTGGCTCAGAGCTGTGACACCGGGTGGCACCAACATCT 756
Db 938 CAGAAGCAGCAGCCAGCCAGCTCAGCTCAGTGTGCTGCTGCTGGGGCCACCATCC 997
QY 757 ATGAGCTCCAGCCCTGGAGAGGCGGGCTGGAGACCCACCATGAGCGCCGCTGGAGG 816
Db 998 ATGCTTGTGATGCTGTGGAGAGTGGGGCTTCCGCTCCCTGCTCATCAACGCTGTGGAGG 1057
QY 817 CTGCCACCTGCCGCGCAAGGAGCT 841
Db 1058 CCTCCTGCATCCGCACACGGGAGCT 1082

RESULT 15
BC022244
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
BC022244 1757 bp mRNA linear PRI 04-FEB-2002
IMAGE:4420238, mRNA, complete cds.
BC022244
BC022244.1 GI:18490812
MGC.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1757)
Strausberg,R.
Direct Submission
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cqabp@remail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK plate: 27 Row: j Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source
1. .1757
/organism="Homo sapiens"
/db_xref="LocusID:5831"
/db_xref="taxon:9606"
/clone="MGC:22061 IMAGE:4420238"
/tissue_type="Lymph, lymphoma"
/clone_lib="NIH_MGC_85"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
406. .921
/codon_start=1

/product="pyrroline-5-carboxylate reductase 1"
/protein_id="AAH22244.1"
/db_xref="GI:18490813"
/translation="MEQLLSVGFCTEVEDLIDAVTGLSGSPAYAFALDALADGG
VKMGLPRLAVRLGAQALLHSEHQQLKDNVSPGATTHALHVLSESGF
RSLILINAVEASCIINTRELQSMADQEQVSPAIAKTKILTIDKVLKLDSPAGTALS
LPRSLAPAGKD"
BASE COUNT 351 a 558 c 505 g 343 t
ORIGIN
Query Match 15.4%; Score 205.4; DB 9; Length 1757;
Best Local Similarity 58.6%; Pred. No. 4.2e-27;
Matches 379; Conservative 0; Mismatches 259; Indels 9; Gaps 1;
QY 204 ACGCACTCCACACAGGAGGTCTGCAGAGCTGCCTGCTGCATCTTTGCCACCAAGCT 263
Db 118 ACACCCCAACAAGGAGACGGTCAGCACAGTATGCTCTTCTGCTGTGAAGCCA 177
QY 264 CATGTGCTGCCAGCTGCTCGGCAGAGGTCTCTGTGGTCCACACTGAACACATCTTG 323
Db 178 CACATCATCCCTTCATCTCGATGAATAGGCGCGACATGAGGACAGACATGTG 237
QY 324 GTGTCCGTGCTGTGGGTCTCTGTAGCACCTCGAGGAGCTGCTGCC----- 374
Db 238 GTGTCTCGCGCGCGCGCTCACCATCAGCTCCATTGAGAAGAAGCTGTCAGCGTTTCGG 297
QY 375 CCAACACACAGGTGCTGCGGGTCTTGCACCACTGCTGTGTGGTCCAGGAAGGGCC 434
Db 298 CCAGCCCCAGGGTCTCCGCTGCATGACCAACACTCCAGTCTGGTGGCGGAGGGGCC 357
QY 435 ATAGTGTGCGCGCGCGCTGCGGAGCAGCGAGACCAAGCTCTCTCAGCATCTG 494
Db 358 ACCGTGTATGCCACAGGACCGCACGCCAGTGGAGGACGGAGGCTCATGGACGCTG 417
QY 495 CTGAGGCGCTGTGGCGGTGTGAGAGTGCCTTGAAGCTTACGTGACATCCACACTGGC 554
Db 418 CTGAGCAGCGTGGGCTTCTGCACGAGGTGGAAGAGGACCTGATTGATGCCGTCCGCGG 477
QY 555 CTCAGTGGCAGTGGCGCTTGTGTGCATCTCCGAGGCGCTGGCTGAAGGAGCC 614
Db 478 CTCAGTGGCAGCGCGCGCTACGATTCACGCCCTGGATGCCCTGGCTGTGGGGGT 537
QY 615 GTCAAGATGGGCATGCCCAGCAGCTTGGCCCAACCGCATCGCTGCCAGACCTCTCTGGG 674
Db 538 GTGAAGATGGGACTTCCAAAGCGCTGCGAGTCCGCTCGGGGCCAGGCCCTCTCGGG 597
QY 675 ACGGCCAAGATGCTGTGTCAGAGGGCCAAACCCAGCCAGCTGGCGTCAAGAGTGTGC 734
Db 598 GCTGCCAAGATGCTGTGCTCACTCAGAACAGCAGCAGCCAGCCAGCTCAAGGACCAAGTCAGC 657
QY 735 ACCCGGGTGCACACCATCTATGGACTTCCACGCCCTGGAGAGCGGGGCTCGGAGCA 794
Db 658 TCTCTGTGGGGCCACCATCCATGCTTGCATGTGTGGAGAGTGGGGGCTTCCGCTCC 717
QY 795 GCCACCATAGCCCGCTGGAGGCTGCCACCTTGGCGGGCCCAAGGAGCT 841
Db 718 CTGCTCATCAACGCTGTGGAGGCTCTCTGTCATCCGCACACGGGAGCT 764

Search completed: July 26, 2003, 06:45:11
Job time : 3572 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:52:06 ; Search time 26 Seconds
(without alignments)
310.072 Million cell updates/sec

Title: US-09-806-536A-14
Perfect score: 1386
Sequence: 1 MAAAEPSRRVGFVGAGRMA.....AATMSAVEAATCRAKELSRK 274

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA:*
- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
 - 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
 - 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
 - 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
 - 5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
 - 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	560.5	40.4	315	3	US-09-099-676-3
2	560.5	40.4	315	4	US-09-565-910-3
3	537.5	38.8	314	3	US-09-099-676-1
4	537.5	38.8	314	4	US-09-565-910-1
5	297.5	21.5	304	1	US-08-665-716-2
6	269	19.4	282	4	US-09-134-001C-3155
7	116	8.4	144	4	US-08-936-165A-445
8	91.5	6.6	436	3	US-08-669-378-4
9	88	6.3	342	4	US-09-364-230-6
10	87.5	6.3	565	3	US-09-008-481A-6
11	87.5	6.3	565	4	US-09-195-666A-5
12	87.5	6.3	565	4	US-09-195-666A-49
13	87.5	6.3	565	4	US-09-309-592-6
14	87.5	6.3	565	4	US-09-635-705-5
15	87.5	6.3	565	4	US-09-635-705-49
16	87.5	6.3	565	4	US-09-634-858A-5
17	87.5	6.3	565	4	US-09-634-858A-49
18	87.5	6.3	1024	4	US-09-562-737-44
19	87	6.3	528	4	US-08-928-213B-8
20	87	6.3	949	4	US-09-196-387-10
21	87	6.3	1327	4	US-09-196-387-2
22	86	6.2	551	3	US-08-886-886-15
23	85.5	6.2	746	2	US-08-785-431-4
24	85.5	6.2	746	4	US-09-205-048-4
25	85.5	6.2	788	2	US-08-785-431-2
26	85.5	6.2	788	4	US-09-205-048-2
27	85	6.1	3729	2	US-08-804-227C-4

28	84.5	6.1	436	3	US-08-669-378-6	Sequence 6, Appli
29	84.5	6.1	436	3	US-08-669-378-10	Sequence 10, Appl
30	84.5	6.1	638	4	US-09-413-814-102	Sequence 102, App
31	83	6.0	2475	4	US-09-413-814-48	Sequence 48, Appl
32	82.5	6.0	333	4	US-09-338-420-4	Sequence 4, Appli
33	82.5	6.0	386	4	US-09-338-420-2	Sequence 2, Appli
34	82.5	6.0	573	3	US-09-008-481A-7	Sequence 7, Appli
35	82.5	6.0	573	4	US-09-195-666A-50	Sequence 50, Appl
36	82.5	6.0	573	4	US-09-309-592-7	Sequence 7, Appli
37	82.5	6.0	573	4	US-09-635-705-50	Sequence 50, Appl
38	82.5	6.0	573	4	US-09-634-858A-50	Sequence 50, Appl
39	82.5	6.0	741	2	US-08-849-480A-9	Sequence 9, Appli
40	82	5.9	5405	4	US-08-718-388-9	Sequence 9, Appli
41	81.5	5.9	436	3	US-08-669-378-2	Sequence 2, Appli
42	81.5	5.9	436	3	US-08-669-378-8	Sequence 8, Appli
43	81	5.8	345	4	US-09-364-230-8	Sequence 8, Appli
44	80	5.8	382	4	US-09-586-719-8	Sequence 8, Appli
45	79.5	5.7	781	4	US-09-766-387-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-099-676-3
; Sequence 3, Application US/09099676
; Patent No. 6100075
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
; TITLE OF INVENTION: HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/099,676
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerione, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0532 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-855-0572
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: 189498

US-09-099-676-3
Query Match 40.4%; Score 560.5; DB 3; Length 315;

[illegible]

RESULT 2
 US-09-565-910-3
 ; Sequence 3, Application US/09565910
 ; Patent No. 6268192
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Corley, Neil C.
 ; APPLICANT: Baughn, Mariah R.
 ; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
 ; TITLE OF INVENTION: HOMOLOG
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows
 ; SOFTWARE: FastSeq for Windows Version 2.0b
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/565,910
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/099,676
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cerrone, Michael C
 ; REGISTRATION NUMBER: 39,132
 ; REFERENCE/DOCKET NUMBER: PF-0532 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-855-0555
 ; TELEFAX: 650-855-0572
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 315 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: GENBANK
 ; CLONE: 189498
 ; US-09-565-910-3

Query Match	40.4%;	Score	560.5;	DB	4;	Length	315;
Best Local Similarity	46.1%;	Pred.	0.1.3e-53;				
Matches	123;	Conservative	46;	Mismatches	87;	Indels	11;
Gaps	5;						
QY	11	VGFVGAGMAGAAGLLRACKVGAQHILASPTDRNLHFOAL---	GCRTTHSNQOEVLQ	67			
DB	3	VGFICAGQL--AFAGFTAAAGVLAHKKIMASSP-DMDLATVSALRKMGVKLLPHNKETVQ	59				
QY	68	SCLLVIFATKPHVLPVLAUEVAPVVTTEHILVSVAAGVSLSTLEELLP---	PNTRLRVLRLV	124			
DB	60	HSVDLVFLAVKPHIITPFIIDEIGADIEDRHVWSCAAGVTTISSEKLSAFRRPAPRVIRM	119				
QY	125	PNLPCVWQEGATVMAWRGHVSSSETKLQHLLEACRCBEVPAYVDIHTGLSSGSAFV	184				
DB	120	TNTPVVVREGATVYATGTHAQVDEGRLMQELLSTVGFCIEVEDLLIDATVGLSSGPAY-	178				
QY	185	CAFSALAEAGVAKMGMPSSLAHRTAATLLGTAKMLHBCQHPAQLRSDVCTPGGTTIYG	244				
DB	179	-AFTALDADGGVKKGLPRPLRAVRIGAAQLLGAAKMLHSEHQPGQIKDNVSSPGGATTHA	237				
QY	245	LHAEQGGGLRAATWSAVEAATCRAKEL	271				
DB	238	LHVLESGGFRSLLINAVEASCIRREL	264				

```

RESULT 3
US-09-099-676-1
; Sequence 1, Application US/09099676
; Patent No. 6100075
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
; TITLE OF INVENTION: HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/099,676
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0532 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-855-0572
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; LIBRARY: PROSNON01
; CLONE: 2278458
; IS-09-099-676-1

```

Query Match.	38.88;	Score	537.5;	DB	4;	Length	314;
Best Local Similarity	45.4%;	Pred. No.	4.4e-51;				
Matches	121;	Conservative	45;	Mismatches	88;	Indels	13;
Gaps	5;						
Qy	11	VGFVWAGRMAGATIAOGLIRAGKVEAQHILASAPTDRLNLCHFQAL	--GCRTHHSNQEVQLQ	67			
Db	3	VGFVGAQQ	----LAYRFTAAGILSAHKIIASSP-EMNLPTVSALRKMGNLRSNKETVK	57			
Qy	68	SCLLVIFATKPHVLPAYLAEVAPVWTEHILVSVAAQVSLSTLEEL	--PPNTRVLRLVL	124			
Db	58	HSDDLFLAVAPHIIPFLDEIGADVQARHIVSCAAGVTISSEVKKLMATQPAKPVIRCM	117				
Qy	125	PNLPVVOEGAIYMARGRHVGSSETKLQHLLEACGCEEVPEAYVDIHTGLSGSVAFV	184				
Db	118	TNTPVVOEGATVYATCTHALVEDQGLLEQLMSSVGFCETEVEEDLIDAVDTGLSGSGPAY	-176				
Qy	185	CAPSEALAECAVKGMPSSLAHRTAAOTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYG	244				
Db	177	-AFWALDADGGKMGFLPRLRIALQIGGAQLLGAAKMLLDSOHFQCLKDNVCSFGGATIIHA	235				
Qy	245	LHLEOQGLRAATMSAYEAATCAKEL	271				
Db	236	LHFUESGGFRSLLINAVEASCIRPEL	262				

RESULTS

```

US-08-665-716-2
; Sequence 2, Application US/08665716
; Patent No. 578922
; GENERAL INFORMATION:
; APPLICANT: KELLY, ROSEMARIE
; APPLICANT: REGISTER, ELIZABETH A
; APPLICANT: MASUREKAR, PRAKASH S
; TITLE OF INVENTION: P5C REDUCTASE GENE FROM ZALERION
; TITLE OF INVENTION: ARBORICOLA
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MERCK & CO., INC.
; STREET: 126 E. LINCOLN AVENUE
; CITY: RAYWAY
; STATE: NEW JERSEY
; COUNTRY: US
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/665,716
; APPLICATION NUMBER: US/08/665,716
; FILING DATE: 23-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: KORSEN, ELLIOTT
; REGISTRATION NUMBER: 32,705
; REFERENCE/DOCKET NUMBER: 19453PV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-594-5493
; TELEFAX: 908-594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 304 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-665-716-2
;
; Query Match 21.5%; Score 297.5; DB 1; Length 304;
; Best Local Similarity 29.3%; Pred No. 1.3e-24;
; Matches 88; Conservative 52; Mismatches 121; Indels 39; Gaps 9

```

[illegible]

Query Match	19.4%;	Score 269;	DB 4;	Length 282;
Best Local Similarity	26.9%;	Pred. No. 1.7e-21;		
Matches	71;	Conservative 54;	Mismatches 137;	Indels 2; Gaps 2
QY	13	FVGAAGMAGATAOGLIRAGKVEAQHI-LASAPTRDRLCHF-QALGCRTHTSNQEVLSQL	70	
Db		: : : : : : : : : : :		
16	FYGAGNMAQAIFTGIINSNNL NANDIYLTNKSEQALKSFAEKLGVNYSYDDDEALLKDAAD	75		
QY	71	LVIATKPCHVLPAVLAEVAPVTEHTLVSVAAAGVSLSLEELLPPNTFVLRVLNPLPCV	130	
Db		: : : : : : : : : : :		
76	VYFGTKPHDPENILANIREHITNDNRFTISIMAGLSIDYIROQLTNPNLPARIMENTNAQ	135		
QY	131	VOEGAIVMARHRHVGSSSETKLLQHLLEACGRCEVEPEAVVDHTGLSGSGVAFVCAFSFA	190	
Db		: : : : : : : : : : : : : :		
136	VGHSVTGISGSNNFDPKSKNEVDELINAFGSVIESEEHLHQVTAITGSGSAFLYHVFEQ	195		
QY	191	LAEGAVKMGPSSSLAHRIAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTTYIGLUHALEQ	250	
Db		: : : : : : : : : : : : : : :		
196	YVKAGTELGLERNQVEESIRNLIICTSKMIERSDLMSQLRKNIITSGKGTTOAGLDALSQ	255		
QY	251	GGLRAATWSAVEATCRAKELSRK	274	
Db		: : : : :		
256	YDIVSMFEEDICIGAVNRSMELSHK	279		

```

> APPLICANT: Black, Michael
> APPLICANT: Burnham, Martin
> APPLICANT: Hodgson, John
> APPLICANT: Knowles, David
> APPLICANT: Lonetto, Michael
> APPLICANT: Nicholas, Richard
> APPLICANT: Pratt, Julie
> APPLICANT: Reichard, Richard
> APPLICANT: Rosenberg, Martin
> APPLICANT: Ward, Judith
> TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
> POLYPEPTIDES AND THEIR USES
> NUMBER OF SEQUENCES: 534
> CORRESPONDENCE ADDRESS:
> ADDRESSEE: SmithKline Beecham Corporation
> STREET: 709 Swedeland Road
> CITY: King of Prussia
> STATE: PA
> COUNTRY: USA
> ZIP: 19406-0939
> COMPUTER READABLE FORM:
> MEDIUM TYPE: Diskette
> COMPUTER: IBM Compatible
> OPERATING SYSTEM: DOS
> SOFTWARE: FastSeq for Windows Version 2.0
> CURRENT APPLICATION DATA:
> APPLICATION NUMBER: US/08/936,165A
> FILING DATE: 24-SEP-1997
> CLASSIFICATION: 536
> PRIOR APPLICATION DATA:
> APPLICATION NUMBER: 60/027,032
> FILING DATE: 24-SEP-1996
> ATTORNEY/AGENT INFORMATION:
> NAME: Gimmi, Edward R
> REGISTRATION NUMBER: 38,891
> REFERENCE/DOCKET NUMBER: P50549
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: 610-270-4478
> TELEFAX: 610-270-5090
>
> INFORMATION FOR SEQ ID NO: 445:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 144 amino acids
> TYPE: amino acid
> STRANDEDNESS: single
> TOPOLOGY: linear
> MOLECULE TYPE: Protein
> US-08-936-165A-445
>
> Query Match 8.4%; Score 116; DB 4; Length 144;
> Best Local Similarity 26.8%; Pred. No. 4.7e-05;
> Matches 30; Conservative 24; Mismatches 56; Indels 2; Gaps 2;
>
QY 16 AGMAGAAAGLIRACKVEAQHI-LASAPTRNLCHF-QALGCRTHSNOEVLOSCLLVI 73
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12 AGNAAQAIPTGIINSSNLDAHDYLNKSNEQAKAFAEKLGYNYSYDDATLLKADAYF 71
QY 74 FATKPHVPLVALEAVPVPTTEHLVSVAAAGVSLSTLEELLPNTVRVLVP 125
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 72 LGTKPHDFDALATRIKPHITKDCXFSIMAGIPTDYIXQOLECONFXARIMP 123
>
> RESULT 8
> US-08-669-378-4
> Sequence 4, Application US/08669378
> Patent No. 6107063
> GENERAL INFORMATION:
> APPLICANT: Mosckel, Bettina
> APPLICANT: Eggeling, Lothar
> APPLICANT: Sahm, Hermann
> TITLE OF INVENTION: PRODUCTION OF L-ISOLEUCINE BY MEANS OF
> RECOMBINANT MICROORGANISMS WITH DEREGULATED THREONINE
> DEHYDRATASE
> TITLE OF INVENTION:

```


NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,981
REFERENCE/DOCKET NUMBER: ARI0-1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-008-481A-6

Query Match 6.3%; Score 87.5; DB 3; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 QGLIRAGKVEAQHILASAPTRNLCHFOALGCRTHSNQEVLOS-CLLVIFATKPHVLP 83
DB 155 ENLLGALLVAGQYIIPEVCLFMNSQLFR--GNRVTKVDSQKFEAFC-----SPNLS- 204
QY 84 VLAEPVPTTEHILVSVAGVSLSTLELLPPNTRVLRVLPNLCV-----QEGA 135
DB 205 -LATVGADVTIADLVKRVNWKDPLVHSHMEHDVALLRLYPASLVRAFLQPLKGV 263
QY 136 IVMARGRHVSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182
DB 264 VLETFGSGNGPSKPDLLQELRAAQRGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAGAVKMGMS--SLAHRTAAQTLLGTAKMLLHGHQHPAQLRSDVCTPGGTT 241
DB 321 LDMTSEAAKLKSLVGLPELSLERR---QELL--AKDLRGEMTLPTADLHQSSPPGSL 375
QY 242 IYGLHAL-----EOGLRAATMSAVEAATCAKEL 271
DB 376 GQGVARLFSFGCQEDSVQDVMPSLALAHAGEL 412

RESULT 11
US-09-195-666A-5
Sequence 5, Application US/09195666A
Patent No. 6140101
GENERAL INFORMATION:
APPLICANT: ARI0, Takeshi
APPLICANT: TANAI, Modaka
APPLICANT: YAMAMOTO, Kozo
APPLICANT: KURIMOTO, Masashi
TITLE OF INVENTION: POLYPEPTIDES HAVING L-ASPARAGINASE ACTIVITY
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 624 Ninth Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/195,666A
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/869,927
FILING DATE: 05-JUN-1997
APPLICATION NUMBER: JP 168,172/96
FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.

REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: ARI0-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-195-666A-5

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 QGLIRAGKVEAQHILASAPTRNLCHFOALGCRTHSNQEVLOS-CLLVIFATKPHVLP 83
DB 155 ENLLGALLVAGQYIIPEVCLFMNSQLFR--GNRVTKVDSQKFEAFC-----SPNLS- 204
QY 84 VLAEPVPTTEHILVSVAGVSLSTLELLPPNTRVLRVLPNLCV-----QEGA 135
DB 205 -LATVGADVTIADLVKRVNWKDPLVHSHMEHDVALLRLYPASLVRAFLQPLKGV 263
QY 136 IVMARGRHVSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182
DB 264 VLETFGSGNGPSKPDLLQELRAAQRGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAGAVKMGMS--SLAHRTAAQTLLGTAKMLLHGHQHPAQLRSDVCTPGGTT 241
DB 321 LDMTSEAAKLKSLVGLPELSLERR---QELL--AKDLRGEMTLPTADLHQSSPPGSL 375
QY 242 IYGLHAL-----EOGLRAATMSAVEAATCAKEL 271
DB 376 GQGVARLFSFGCQEDSVQDVMPSLALAHAGEL 412

RESULT 12
US-09-195-666A-49
Sequence 49, Application US/09195666A
Patent No. 6140101
GENERAL INFORMATION:
APPLICANT: ARI0, Takeshi
APPLICANT: TANAI, Modaka
APPLICANT: YAMAMOTO, Kozo
APPLICANT: KURIMOTO, Masashi
TITLE OF INVENTION: POLYPEPTIDES HAVING L-ASPARAGINASE ACTIVITY
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 624 Ninth Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/195,666A
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/869,927
FILING DATE: 05-JUN-1997
APPLICATION NUMBER: JP 168,172/96
FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: ARI0-2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)628-5197
TELEFAX: (202)737-3528
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-195-666A-49

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 OGLIRAGKVEAQHILASATDRNLCHFQALGCRTHSHNOEVLQS-CLLVIFATKPHVLP 83
Db 155 ENLLGALLVAGQYIIPEVCLFMNSQLFR--GNRVTKVDSQKFEAFC-----SPNLSP- 204
QY 84 VLAEPVAVVTTEHILSVAAAGVSLSTLELLPENTRLVRLVLPNLPVW-----QEQA 135
Db 205 -LATVGADVTIAMDVLVRKVNWKDPLVVHSMEDVALLRLYPCIPASIVRAFLQPLKGV 263
QY 136 IVMARGRHVGSSTKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSGV--A 182
Db 264 VLETFGSGNGPSKPDLLQELRAAQRGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAEAGVAKMGMP--SLAHRIAQTLLGTAKMLLHGHQHPAQLRSDVCTPGGTT 241
Db 321 LDMTSEAALAKLSYVLGLPELSLERR---QELL--AKDLRGEMTLPTADLHQSPPPGSTL 375
QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271
Db 376 GOGVARLFSLFCQEDSVQDAVMPSLALAHAGEL 412

RESULT 13
US-09-309-592-6
Sequence 6, Application US/09309592
Patent No. 6274367
GENERAL INFORMATION:
APPLICANT: ARIO, Takeshi
APPLICANT: TANIAI, Madoka
APPLICANT: TORIGOE, Kakuji
APPLICANT: KURIMOTO, Masashi
TITLE OF INVENTION: DNA CODING FOR MAMMALIAN L-
TITLE OF INVENTION: ASPARAGINASE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/309,592
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/008,481
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 95-42564
FILING DATE: 08-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,981

REFERENCE/DOCKET NUMBER: ARIO-1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-309-592-6

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 OGLIRAGKVEAQHILASATDRNLCHFQALGCRTHSHNOEVLQS-CLLVIFATKPHVLP 83
Db 155 ENLLGALLVAGQYIIPEVCLFMNSQLFR--GNRVTKVDSQKFEAFC-----SPNLSP- 204
QY 84 VLAEPVAVVTTEHILSVAAAGVSLSTLELLPENTRLVRLVLPNLPVW-----QEQA 135
Db 205 -LATVGADVTIAMDVLVRKVNWKDPLVVHSMEDVALLRLYPCIPASIVRAFLQPLKGV 263
QY 136 IVMARGRHVGSSTKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSGV--A 182
Db 264 VLETFGSGNGPSKPDLLQELRAAQRGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAEAGVAKMGMP--SLAHRIAQTLLGTAKMLLHGHQHPAQLRSDVCTPGGTT 241
Db 321 LDMTSEAALAKLSYVLGLPELSLERR---QELL--AKDLRGEMTLPTADLHQSPPPGSTL 375
QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271
Db 376 GOGVARLFSLFCQEDSVQDAVMPSLALAHAGEL 412

RESULT 14
US-09-635-705-5
Sequence 5, Application US/09635705
Patent No. 6368845
GENERAL INFORMATION:
APPLICANT: ARIO, Takeshi
APPLICANT: TANIAI, Madoka
APPLICANT: YAMAMOTO, Kozo
APPLICANT: KURIMOTO, Masashi
TITLE OF INVENTION: POLYPEPTIDES HAVING L-ASPARAGINASE ACTIVITY
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 624 Ninth Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/635,705
FILING DATE: 09-Aug-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/195,666
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 168,172/96
FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: ARIO=2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)628-5197
TELEFAX: (202)737-3528
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-635-705-5

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 QGLIRAGKVEAQHILASAPDRNLCHFOALGCRTHSNQEVLOS-CLLVIFATKPHVLP 83
Db 155 ENLIGALLVAGQYIPEVCLFMNSQLFR--GNRVTKVDSQKFEAF-----SPNLS- 204
QY 84 VLAEPVAVVTEHILSVSAAGVSLSTLEELLPPNTRVLRVLPNLPV-----QEGA 135
Db 205 -LATVGADVTIANDLVKRVNWKDPLVVHNSMEHDVALLRYPASLVRALQPLKGV 263
QY 136 IVMARGHVSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182
Db 264 VLETFGSGNGPSKPDLLQELRAAQRGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAEAGVAKMGMP--SLAHRIAQTLLGTAKMLLHGOHPAQLRSDVCTPGGTT 241
Db 321 LDMTSEALAKLSVYGLPELSLERR--QELL--AKDLRGEMTLPTADLHOSPPGSTL 375
QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271
Db 376 GOGVARLFLSGCOEDSVQDVMPSLALALAHAGEL 412

RESULT 15

US-09-635-705-49
Sequence 49, Application US/09635705
Patent No. 6368845
GENERAL INFORMATION:

APPLICANT: ARIO, Takeshi
TANAI, Modaka
YAMAMOTO, Kozo
KURIMOTO, Masashi

TITLE OF INVENTION: POLYPEPTIDES HAVING L-ASPARAGINASE ACTIVITY
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 624 Ninth Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/635,705
FILING DATE: 09-Aug-2000

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/195,666
FILING DATE: <Unknown>
APPLICATION NUMBER: JP 168,172/96

FILING DATE: 07-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: ARIO=2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)628-5197
TELEFAX: (202)737-3528
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 49:
US-09-635-705-49

Query Match 6.3%; Score 87.5; DB 4; Length 565;
Best Local Similarity 24.5%; Pred. No. 0.52;
Matches 68; Conservative 36; Mismatches 124; Indels 49; Gaps 13;
QY 25 QGLIRAGKVEAQHILASAPDRNLCHFOALGCRTHSNQEVLOS-CLLVIFATKPHVLP 83
Db 155 ENLIGALLVAGQYIPEVCLFMNSQLFR--GNRVTKVDSQKFEAF-----SPNLS- 204
QY 84 VLAEPVAVVTEHILSVSAAGVSLSTLEELLPPNTRVLRVLPNLPV-----QEGA 135
Db 205 -LATVGADVTIANDLVKRVNWKDPLVVHNSMEHDVALLRYPASLVRALQPLKGV 263
QY 136 IVMARGHVSSETKLLQHLLEA-----CGRC--EEVPEAYVDIHTGLSGSV--A 182
Db 264 VLETFGSGNGPSKPDLLQELRAAQRGLIMVNCSQLRGSVTPGYA---TSLAGANIVSG 320
QY 183 FVCAFSEALAEAGVAKMGMP--SLAHRIAQTLLGTAKMLLHGOHPAQLRSDVCTPGGTT 241
Db 321 LDMTSEALAKLSVYGLPELSLERR--QELL--AKDLRGEMTLPTADLHOSPPGSTL 375
QY 242 IYGLHAL-----EOGGLRAATMSAVEAATCRAKEL 271
Db 376 GOGVARLFLSGCOEDSVQDVMPSLALALAHAGEL 412

Search completed: July 21, 2003, 09:56:32
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 05:39:44 ; Search time 335 Seconds
(without alignments)
8239.687 Million cell updates/sec

Title: US-09-806-536A-29
Perfect score: 1338
Sequence: 1 ggtgagcagctctgtccga.....ggggcagcagcatatggggt 1338

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1307	97.7	2331	15	US-10-161-418A-2
2	1127.8	84.3	1178	15	US-10-161-418A-7
3	857	64.1	860	15	US-10-161-418A-6
4	226.4	16.9	1848	15	US-10-161-418A-3
5	224.8	16.8	1792	11	US-09-880-107-2400
6	224.8	16.8	1792	15	US-10-161-418A-1
7	214.4	16.0	1715	15	US-10-161-418A-5
8	211.2	15.8	1028	15	US-10-161-418A-4
9	203.4	15.2	1478	11	US-09-925-300-278
10	201.8	15.1	1708	15	US-10-161-418A-8
11	201.8	15.1	1742	10	US-09-912-717-2
12	167	12.5	999	15	US-10-161-418A-9
13	156.4	11.7	807	15	US-10-156-761-4708
14	156.4	11.7	9025608	15	US-10-156-761-1
15	129	9.6	454	12	US-09-918-995-11309
16	118.6	8.9	371	12	US-09-918-995-181

17	115.6	8.6	482	12	US-09-918-995-31035
18	84.4	6.3	810	11	US-09-738-626-460
19	84.4	6.3	3309400	11	US-09-738-626-1
20	83.2	6.2	415	12	US-09-918-995-17742
21	73.4	5.5	888	11	US-09-726-397A-4
22	73.4	5.5	888	11	US-09-712-363-28
23	73.4	5.5	86114	14	US-10-080-170-648
24	72.8	5.4	931	15	US-10-027-632-161469
25	72.8	5.4	931	15	US-10-027-632-161470
26	70.6	5.3	780	11	US-09-974-300-343
27	69.6	5.2	843	15	US-10-027-632-170626
28	62	4.6	855	15	US-10-128-714-2252
29	62	4.6	855	15	US-10-128-714-2252
30	59.8	4.5	1014	11	US-09-070-927A-157
31	54.6	4.1	2908	15	US-10-128-714-252
32	54.6	4.1	2908	15	US-10-128-714-252
33	54.4	4.1	908	15	US-10-128-714-1252
34	54.4	4.1	908	15	US-10-128-714-1252
35	53.6	4.0	2209	11	US-09-726-397A-2
36	52.2	3.9	334	11	US-09-796-692-6261
37	52.2	3.9	334	15	US-10-040-862-6261
38	48.2	3.6	2817	15	US-10-156-761-390
39	47.4	3.5	888	15	US-10-156-761-1531
40	47.4	3.5	9025608	15	US-10-156-761-1
41	47.2	3.5	795	15	US-10-156-761-1729
42	46.4	3.5	1359	15	US-10-156-761-6088
43	45.6	3.4	834	15	US-10-156-761-1113
44	44.4	3.3	903	15	US-10-156-761-2474
45	44.4	3.3	1645	15	US-10-037-270-807

ALIGNMENTS

RESULT 1

US-10-161-418A-2
; Sequence 2, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161.418A
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 2331
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-161-418A-2

Query Match	97.7%	Score 1307	DB 15	Length 2331
Best Local Similarity	99.2%	Pred. No. 6.3e-309		
Matches 1313	Conservative	0	Mismatches 10	Indels 0
Gaps	0			
QY	15	GTCCGAGGCAACAGATGCGAGCTGCGGCGCGCTTCGCGCGCGTGGGCTTCGTGGGC	74	
Db	12	GTCCGAGGCAACAGATGCGAGCTGCGGCGCGCTTCGCGCGCGTGGGCTTCGTGGGC	71	
QY	75	CGGGCCCGCATGGCGGGGCCATCGCGAGGCGCTCATCAGACGAGGAGAAAGTGAAGCT	134	
Db	72	CGGGCCCGCATGGCGGGGCCATCGCGAGGCGCTCATCAGACGAGGAGAAAGTGAAGCT	131	
QY	135	CAGCACATCTGGCCAGTGCACCAACAGACAGGAACTATGTCACTTTCAAGCTCTGGGT	194	
Db	132	CAGCACATCTGGCCAGTGCACCAACAGACAGGAACTATGTCACTTTCAAGCTCTGGGT	191	
QY	195	TCCCGGACACGACACTCCCAACAGGAGGTGTGACAGAGCTGCCTGCTCGTCTCTTCC	254	


```
Db 791 CTCCTGCATCCGACACGGGAGCT 814
RESULT 6
US-10-161-418A-1
; Sequence 1, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: PSYCRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1792
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-161-418A-1
Query Match 16.8%; Score 224.8; DB 15; Length 1792;
Best Local Similarity 58.0%; Pred. No. 1.6e-45;
Matches 466; Conservative 0; Mismatches 317; Indels 21; Gaps 3;
QY 56 GCGCGTGGCTTCGTGGCGGGGGCCGATGGCGGGGGCCATCGCGAGGCGCTCATCAG 115
Db 14 GAGCGTGGCTTCATCGCGGCTGGCCAGCTGGCTTTCCTCCCTGGCCAGGCGCTTCACAGC 73
QY 116 AGCAGAGAAAGTGAAGTCTAGCAGCATACTGGCCAGTGCACCAACAGACAGAGAACCTATG 175
Db 74 AGCAGGCGTCTGGCTGCCACAGATATAGCTAGCTCC--CAGACATGGACCTGGC 130
QY 176 TCACATTCAGCTC-----TGGGTTGCCGACGAGCAGCTCCACACAGGAGGCT 226
Db 131 CACAGTTCTGCTCAGGAAGATGGGGGTGAAGTTGACACCCCAACAAAGGAGACGGT 190
QY 227 GCAGAGCTGCTGCTGCTCATCTTTGCCACCAAGCCTCATGTGCTGCCAGCTGTCTGGC 286
Db 191 GCAGCACAGTATGATGCTCTTCTGCTGTGAAGCCACACATCATCTCCCTTCATCTCGA 250
QY 287 AGAGTGGCTCTGTGGTCCACCATCTGACACATCTTGTGTCCGTGGCTGTGGGGTGC 346
Db 251 TGAATATAGGCGCCGACATTTGAGGACAGACACATTTGTTGTTGCTCGCGGGCGCGGTAC 310
QY 347 TCTGAGCACCTGGAGGAGCTGCTG-----CCCCCAACACACAGGCTGCTCGGGT 397
Db 311 CATCAGCTCCATTGAGAGAAGCTGTACAGCGTTTCGGCCAGCCCCCAGGGTCACTCCGCTG 370
QY 398 CTTGCCCAACCTGCCCTGTGGTCCAGGAAGGGCCATAGTGTATGGCGGGGGCGGCCA 457
Db 371 CATGACCAACTCCAGTCTGCTGCGGGAGGGGCCACCGGTATGCCACAGGACGCA 430
QY 458 CGTGGGAGCAGGAGACCAACTCTGCAGCATCTGTGAGGCGCTGTGGCGGTGTA 517
Db 431 CGCCCCAGGTGGAGAGCGGAGGCTCATGGAGCAGCTGTGAGCAGGTGGGCTTCTGCAC 490
QY 518 GGAGGTGCTGAAGCCTACGTCACATCCACATCGCCTCAGTGGCAGTGGCGTGGCCTT 577
Db 491 GGAGGTGGAAGAGGACCTGATTGATGCCGTACAGGGGCTCAGTGGCAGGGCGGCCCTA 550
QY 578 CTTGTGTGATTTCTCGAGGCCCTGGCTGAAGAGCCGCTCAAGATGGGATGCCAGCAG 637
Db 551 CGCATTCACAGCCCTGGATGCCCTGGCTGATGGGGTGTGAAGATGGGACTTCCAAGGCG 610
QY 638 CTTGGCCCAACCGATCGCTGCCACACCTGTCTGGGAGCGGCAAGATGCTGCTGCACGA 697
Db 611 CTTGGCAGTCCGCTCGGGGGCCAGGCCCTCTCTGGGGGCTGCCAAGATGCTCTGCACCTC 670
```

```
QY 698 GGCCCAACCCAGCCAGCTCGCTCAGACGTTGTGCACCCGGGTGGCACCACTCATTA 757
Db 671 AGAACAGCACCCAGGCCAGCTCAAGGACACACGTCAGCTCTCTGTGGGGCCACCATCA 730
QY 758 TGGACTCCAGCCCTGGAGCAGGCGGCTGGCAGCAGCCACCATGAGGCCCTGGAGGC 817
Db 731 TGCCTTGCATGTCTGGAGAGTGGGGCTTCCGCTCCCTGCTCATCAACGCTGTGGAGGC 790
QY 818 TGGCACTTGGCGGGCCAAAGAGCT 841
Db 791 CTCCTGCATCCGACACGGGAGCT 814
RESULT 7
US-10-161-418A-5
; Sequence 5, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: PSYCRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1715
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-161-418A-5
Query Match 16.0%; Score 214.4; DB 15; Length 1715;
Best Local Similarity 58.0%; Pred. No. 5.4e-43;
Matches 466; Conservative 0; Mismatches 316; Indels 22; Gaps 4;
QY 56 GCGCGTGGCTTCGTGGCGGGGGCCGATGGCGGGGGCCATCGCGAGGCGCTCATCAG 115
Db 281 GAGCGTGGCTTCATCGCGCTGGCCAGCTGGCTTTTGGCCCTGGCCAAAGGCTTTACAGC 340
QY 116 AGCAGAGAAAGTGAAGCTCAGACATACTGGCCAGTGCACCAACAGACAGAGAACCTATG 175
Db 341 AGCAGGCGTCTTGGCTGCCACACAGATATAGCTAGCTCC--CAGACATGGACCTGGC 397
QY 176 TCACATTCAGCTC-----TGGGTTGCCGAGCAGCAGCTCCACACAGGAGGCT 226
Db 398 CACAGTTTCTGCTCAGGAAGATGGGGTGAAGTTGACACCCCAACAAAGGAGACGGT 457
QY 227 GCAGAGCTGCTGCTGCTCATCTTTGCCACCAAGCCTCATGTGCTGCCAGCTGTCTGGC 286
Db 458 GCAGCAGTGTATGTGCTCTTCTGGCTGTGAAGCCACACATCATCCCTTCATCTCGGA 517
QY 287 AGAGTGGCTCCTGTGCTCACCCTGAGCAGCATCTTGGTGTCCGTGGCTGTGGGGTGC 346
Db 518 TGAATATAGGCGCCGACATTTGAGGACAGACACATTTGTGTGCTCGCGGGCGCGCTAC 577
QY 347 TCTGAGCACCTGGAGAGCTGCTG-----CCCCCAACACACAGGCTGCTGGCGGT 397
Db 578 CATCAGCTCCATTGAGAAGAGCTGTGAGCGCTTTCGGCCAGCCCCCAGGGTCACTCCGCTG 637
QY 398 CTTGCCCAACCTGCCCTGTGGTCCAGGAAGGGCCATAGTGTATGGCGGGGGCGGCCA 457
Db 638 CATGACCAACTCCAGTCTGGTGGGGGGGCCACCGGTATGCCACAGAGCCAC 697
QY 458 CGTGGGAGCAGCAGAGACAAAGCTCTGCAGCATCTGTGGAGGCGCTGTGGCGGTGTA 517
Db 698 GCCCAGGTGGAG-GACGGGAGGCTCATGGAGCAGCTGTGAGCAGCGTGGGCTTCTGCAC 756
QY 518 GGAGTGGCTGAAGCCTACGTCAGCATCCACATCTGGGCTCAGTGGGAGTGGCGGT 577
```

Db 757 GGAGGTGGAGAGGACCTGATTGATGCCCTACGGGGCTCAGTGGCAGCGGCCCGCCTA 816
QY 578 CBTGTGTGATTTCTCCAGGCGCTTGGCTGAAGAGCCGTCGAAGATGGGCATGCCAGCAG 637
Db 817 CGCATTCACAGCCCTGATGCGCTGGCTGATGGGGTGTGAAGATGGGACTTCCAAGCGG 876
QY 638 CTTGGCCACACCGATCGCTGCCAGACCCCTGTGGGGAGGCCAAGATGCTGCTGCACGA 697
Db 877 CTTGGCAGTCCGCTCGGGGCCAGAGCCCTCTCTGGGGGTGCAAGATGCTGCTGCACCTC 936
QY 698 GGCCCAACACCCAGCCAGCTGGCTCAGACCTGTGCACCCGGGTGGCCACCAACATCTA 757
Db 937 AGACAGACACCCAGCCAGCTCAAGACAACTCAGCTCTCTGTGGGGCCACCATCCA 996
QY 758 TGGACTCCAGCCCTGGAGCAGGGGGGTGGAGCAGCCACCATGAGCGCGTGGAGGC 817
Db 997 TGCCTTGCATGTGCTGGAGAGTGGGGCTTCCGCTCCCTGCTCATCAACGCTGTGGAGGC 1056
QY 818 TGGCAGCTCCGGGCCAAGGAGCT 841
Db 1057 CTCTGCTATCCGCACACGGGAGCT 1080

RESULT 8

US-10-161-418A-4
; Sequence 4, Application US/10161418A
; Publication No. US20030036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1028
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (999)..(999)
; OTHER INFORMATION: "n" is A, C, G, or T
US-10-161-418A-4

Query Match 15.8%; Score 211.2; DB 15; Length 1028;
Best Local Similarity 58.2%; Pred. No. 3.3e-42;
Matches 438; Conservative 0; Mismatches 293; Indels 21; Gaps 3;
QY 56 GCGCTGGGCTTCGTGGCGCGGGCGCATGCGGGGGCCATCGCGAGGCGCTCATCAG 115
Db 61 GAGCTGGGCTTCATCGCGCTGGCCAGCTGGCTTTTGGCCCTGGCCAGGCTTCACAGC 120
QY 116 AGCAGAAAGTGGAGCTCAGCATACTGGCCAGTGCACCAACAGACAGAACTATG 175
Db 121 AGCAGGCTCTTGGTGGTCCCAAGATAATGGCTAGCTCC--CAGACATGGACCTGGC 177
QY 176 TCACCTTTCAGCTC-----TGGGTTGCCGACACGACATCCCAACAGGAGGTGCT 226
Db 178 CACAGTTTCTGCTCAGGAAGATGGGGTGAAGTTGACACCCCAACAGGAGAGCGGT 237
QY 227 GCAGAGCTGCTGCTGCTATCTTTGGCCACCAACCTCATGTGCTGCCAGCTGTCCTGGC 286
Db 238 GCAGCAGAGTGTATGTGCTCTCTCTGGTGTGAAGCCACACATATCCCTTCATCCTGGA 297
QY 287 AGAGTGTGCTGCTGCTGCTACCATGACACATCTTGGTTCCTGCTGCTGGGTGCTC 346
Db 298 TGAATAGCGCCGACATTTAGGAGACAGACATTTGTGTTGCTGCTGGCGCGCGCGCTC 357

QY 347 TCTGAGCACCCTGTGAGGAGCTGCTG-----CCCCCAACACACGGGTGCTGGGGT 397
Db 358 CATCAGCTCCATTCAGAAGAAGCTGTGAGCGTTTCGGCCAGCCCGCAGGCTCATCGCGTG 417
QY 398 CTTGCCCAACCTGCCCTGTGTGTCCAGGAAGAGGGGCCATAGTATGATGCGCGGGCGGCCA 457
Db 418 CATGACCAACACTCAGTCTGTGTGCGGAGGGGGCCACCGTGTATGCCACAGCAGCA 477
QY 458 CGTGGGAGCAGCAGACCAAGCTCCCTGCAGCATCTGCTGGAGCCCTGTGGCGGTGTGA 517
Db 478 CGCCAGGTGGAGCGGGAGGCTCATGGAGCAGCTGCTGAGCAGCGTGGGCTTCTGCAC 537
QY 518 GGAGTGCCTGAAGCCCTACGTCGACATCCACACTGGCCTCAGTGGAGTGGCGTGGCCTT 577
Db 538 GGAGTGAAGAGGACCTGATTGATGCCGTACGGGGCTCAGTGGCAGCGGCCCGCCTA 597
QY 578 CGTGTGTGATTTCCAGGCGCTGCTGAAGAGCGGCTCAAGATGGGCAATGCCAGCAG 637
Db 598 CGCATTCACAGCCCTGGATGCCCTGCTGATGGGGTGTGAAGATGGGACTTCCAAGCG 657
QY 638 CTTGGCCACCGCATCGCTGCCAGACCCCTGCTGGGACGGCCCAAGATGCTGTGCACGA 697
Db 658 CTTGGCAGTCCGCTCGGGCCCGCAGGCCCTCTCGGGGCTGCCAAGATGCTGTGCATC 717
QY 698 GGGCCAAACACCCAGCCAGCTGCGCTCAGACGTGTGCACCCGGGTGGCACCACCATCTA 757
Db 718 AGAAGCAGCACCAGCCAGCTCAAGGACAACAGCTCTCCTGTGGTGGGGGACCATCCA 777
QY 758 TGGACTCCAGCCCTGGAGCAGGCGGGCTGC 789
Db 778 TGCTTGCATGTGCTGGAGAAAGTGGGGGCTTC 809

RESULT 9

US-09-925-300-278
; Sequence 278, Application US/09925300
; Patent No. US20020151881A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 278
; LENGTH: 1478
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-278

Query Match 15.2%; Score 203.4; DB 11; Length 1478;
Best Local Similarity 57.9%; Pred. No. 2.6e-40;
Matches 466; Conservative 0; Mismatches 316; Indels 23; Gaps 5;
QY 56 GCGCGTGGCTGCTGGCGCGGGCGCATGCGGGGGCCATCGCGAGGCGCTCATCAG 115
Db 78 GAGCGTGGGCTTCATCGGCGCTGGCCAGCTGGCTTTTGGCCCTGGCCAGGGGCTGCACAG 137
QY 116 -AGCAGAAAGTGGAGCTCAGCATACTGGCCAGTGCACCAACAGACAGAACTAT 174
Db 138 CAGCAGCGCTCTGGCTGCCCAAGATAATGGCTAGCTCCC--CAGACATGGACCTGG 194
QY 175 GTCACCTTCAAGCTC-----TGGGTTGGCGGACACGACATCCCAACAGGAGGTGC 225
Db 195 CCACAGTTTCTGCTCTCAGGAAGATGGGGTGAAGTTGACACCCCAACAGAGAGCGG 254
QY 226 TGCAGAGCTGCCTGCTCGTCACTCTTTGCCACCAAGCCTCATGTGCTGCCAGCTGCTCTGG 285

Db 255 TGCAGCAGTGTGTCCTCTGGCTGTGAAG-CACACATCATCCCCCTTCATCTGG 313
Qy 286 CAGAGGTGGCTCTGTGTGTACCACTTGAACACATCTTGGTGTGGTGTGGTGTGGTGT 345
Db 314 ATGAATAGGCGCCGACATTGAGNACAGACATTTGGTGTCTGTGGCGCGCGGTCA 373
Qy 346 CTCTGAGCACCCTGAGAGGTGTGTGCCC-----CNAACACAGGTGTGTGGGG 396
Db 374 CCATCAGCTCCATTGAGAAGAGCTGTGAGCGCTTTTGGGCCAGCGCCAGGGGTCTACGGCT 433
Qy 397 TCTTCCCAACCTGCTGTGTGTGTCAGGAAGGGCCATAGTGTATGCGCGGGCGCGCC 456
Db 434 GCATGACCAACTCTCACTGTGTGTGTCAGGAAGGGCCATAGTGTATGCGCGGGCGCGCC 493
Qy 457 AGCTGGGAGCAGCAGACCAAGCTCTCTGACGATCTGTCTGAGGCGCTGTGGGCGGTGTG 516
Db 494 AGCGCCAGGTGGAGCAGCGGAGGCTCATGTGAGCAGCTGCTGAGCAGCGTGGCTTCTGCA 553
Qy 517 AGGAGGTGCTTGAACCTTACGTGACATCCACACTGGCTCAGTGGCGAGTGGCGTGGCT 576
Db 554 CGGAGGTGGAAGAGACCTGTATGATCCGCTCAGGGGCTCAGTGGCAGCGCGCGCGCT 613
Qy 577 TCGTGTGTGCTATCTCCGAGGCGCTGTGCTGAAGGAGCGCTCAAGATGGGCATGCCAGCA 636
Db 614 AGCATTCACAGCCCTGTGATGCCCTGTGATGGGGTGTGAGATGGGACTTCAAGGC 673
Qy 637 GCCTGGCCCAACCGCATGCTGCCCCAGACCTGTCTGGGAGCGGCCAAGATGTGTGACG 696
Db 674 GCCTGGCAGTCCGCTCGGGGCCAGAGCGCTCTCTGGGGCTGCCAAGATGTGTGCACT 733
Qy 697 AGGGCCACACACCCAGCCAGCTGCGCTCAGAGCTGTGCACCCCGGTGGCACCACCATCT 756
Db 734 CAGAACAGCACCAGCGCCAGCTCAAGAGCAAGCTCAGCTCTCTGGTGGGCGCCATCC 793
Qy 757 ATGAGCTCACCCCTCGAGCAGGGGCGCTCGAGCAGCCACCATGAGCGCGCTGTGAGG 816
Db 794 ATGCTTGTGATGTGTGAGAGTGGGGCTTCCGCTCCCTGCTCATCAACGCTGTGGAG 853
Qy 817 GTGGCAGCTGCGGGCGCAAGGAGCT 841
Db 854 CCTCCTGCTATCGGCACACGGAGCT 878

RESULT 10
US-10-161-418A-8
; Sequence 8, Application US/10161418A
; Publication No. US2003036078A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: P5CRs AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-055
; CURRENT APPLICATION NUMBER: US/10/161,418A
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: US 60/296,080
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,509
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1708
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-161-418A-8

Query Match 15.1%; Score 201.8; DB 15; Length 1708;
Best Local Similarity 55.0%; Pred. No. 6.3e-40;
Matches 448; Conservative 0; Mismatches 352; Indels 15; Gaps 2;
Qy 42 GAGCCGTCTCCGGCGCTGTGGCTTTCGTGGCGGGCGCGCATGGCGGGCGCATCGCG 101
Db 144 GAACCGGGACCATGAGCTGGGCTTTCATCGGGCGCGCGCATGGCGCTATGCTGTGGCG 203

Qy 102 CAGGGCTCATCAGACGAGAAAGTGGAAAGCTCAGCACAATCTGCGCAGTGCACCAACA 161
Db 204 CGGGGCTTCACGGCCGAGGACTCTGTGGCTCACAAGATAATAGCAGTCCCCAGAA 263
Qy 162 GACAGGAACG-----TATGTCACTTTCAAGCTCTGGGTTCGCGGACCGCAGCTCCAAC 215
Db 264 ATGAACCTGCCAGGTGTCCGCGCTCAGGAAGATGGTGTGAACCTGACACGACGAC 323
Qy 216 CAGGAGTGTGTGAGAGTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 275
Db 324 AAGGAGAGCGTGAAGCAGACGAGCTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 383
Qy 276 GCTGTCTGTGACAGGT 335
Db 384 TTATCTCTGTGATGAGTTGGGCGGAGCTCAAGACCAAGACATCTGTGTGTGTGTGTGT 443
Qy 336 GCTGGGTGTCTGTGAGCAGCCCTGGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 386
Db 444 GCTGGTGTCAACATCAGCTCTGTGGAGAAGAGCTGATGCATTTCCAGCCAGCCCAAA 503
Qy 387 GTGTGTGGGT 446
Db 504 GTGATTCGCTGCATGACCAACACACACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 563
Qy 447 CGGGGCGCGCAGCTGGGGAGCAGCAGACCAAGCTCTGTGTGTGTGTGTGTGTGTGTGT 506
Db 564 ACGGGCACCATATGCCCTGTGTGGAGATGGGAGCTCTGTGTGTGTGTGTGTGTGTGT 623
Qy 507 GGGCGGTGTGAGGAGTGCCTGAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 566
Db 624 GGTCTTGTGCTGAGGTGGAAGAGAGCTCTCATGTATGCTGTGTGTGTGTGTGTGTGT 683
Qy 567 GCGTGTGCTGT 626
Db 684 GGGCTGCTGT 743
Qy 627 ATGCCAGCAGCTGGCGCCACCGCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 686
Db 744 TTGCCAGCGCTGTGCAATCCAACTCGGGGCCAGGCTTGTGTGTGTGTGTGTGTGTGT 803
Qy 687 CTGTGTACGAGGCGCCAAACAGCCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 746
Db 804 CTGTGTGACTCGAGCAGCATCCATGCGCTTAAAGACATGTCTGTGTGTGTGTGTGT 863
Qy 747 ACCACCATCTATGACTCCAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 806
Db 864 GCCACCATCCAGCGCTGT 923
Qy 807 GCGCTGGAGGCTGCACTGCGGGGCCAAGGAGCT 841
Db 924 GCAGTTGAGGCTCTGTATCCGAACAGAGAGCT 958

RESULT 11
US-09-912-717-2
; Sequence 2, Application US/09912717
; Patent No. US20020081691A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; Corley, Neil C.
; Baughn, Mariah R.
; TITLE OF INVENTION: DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

Db	720	GGCGCTGCCTATGCATTATCATGCTCTGGACGATTTGGCTGATGCTGGGCTGAAGATGGGT	779
Qy	627	ATGCCACGACGCTGGCCACCGATCTGCTGCCAGACCTGCTGGGACGGCCAAAGTG	686
Db	780	TTGGCACGGCGCTGGCAATCCAACCTGGGGCCACAGCTTTGCTGGGAGCTGCCAAGTG	839
Qy	687	CTGCTTCACGAGGGCCAAACCCAGCCCAAGTCGGCTCAGACGTGTGCACCCCGGTGGC	746
Db	840	CTGCTGAGCTCGGACAGCATCCATGCCAGCTTAAGGACAAATCTCTGCTCCCTGGGGGA	899
Qy	747	ACCACCATATGTGAGCTCCAGCCCTGGAGCAGGGCGGCTGGAGCGCCACCACTGAGC	806
Db	900	GCCACCATCCAGCCCTGCATTTCTAGAGAGTGGGGGCTTCCGCTCTCTGCTCATCAAT	959
Qy	807	GCGGTGGAGGCTGCCACCTGCCGGGGCCAAAGAGCT	841
Db	960	GCAGTTTGAGGCTCTCTGTATCCGAACACGAGAGCT	994

FILING DATE: <Unknown>

NAME: CERRONE, MICHAEL C
REGISTRATION NUMBER: 39-132

TELECOMMUNICATION INFORMATION:

TELEFAX: 650-855-0572
TELEX: <Unknown>

SEQUENCE CHARACTERISTICS

TYPE: nucleic acid
STRANDEDNESS: single

IMMEDIATE SOURCE:

CLONE: 2278458
SEQUENCE DESCRIPTION

Local Similarity 35.0%; Pred. No: 6.3e-40;
448: Conservative 0; Mismatches 353;
Total: 15 Cons

42 GAGCCGTCCTCCGCGCGCGTGGGCTTCGTGGCGCGGGCCGCATGGCGGGGCATCGCG 101

180 GAACCGCGACCAATGAGCGTGGGCTTCATCGGGGGCCGGCCAGCTGGCCTATGCTCTGGCG 239

[illegible]

162 GACAGGACC-----TATCTCATCTTCACCTCCCTCCTCCCCTCCCC

300 ATGAACCTGCCACGGTGTCCGGCTCAGGAAGATGGGTGTAACCTGACACGCAGCAAC 359

216 CAGGAGGTGCTGCAGAGCTGCCCTGCTCGTCATCTTGCCACCAAGCCTCATGTGCTGCCA 275

[illegible]

11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 10

336 GCTGGGTGTCTGTAGCACCCCTGGAGGAGCTGCTG-----CCCCCAACACACGG 386

480 GCTGGTGTCAACCATCAGCTCTGTGGAGAAGAGCTGATGGCATTCAGCCAGCCCCCAA 539

CCG CAGGAAAGGGCCATAGTGTGGCG 446

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

600 ACGGCACCCATGCCCTGGTGGAGGATGGGACGTCATCAGCACCTC 650

507 GGGCGGTGTGAGGAGTGCCCTGAAGCCTACGTCGACATCCACACTGGCCTCAGTGGCAGT 566

000 GGCATTCTGCACTGAGGGTGGAAAGAGGACCTCATTCGATGCCGTCACGGGGCTCAGTGGCACC 719

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84


```
QY 436 TAGTGATGGCGGGGGCCGCCACGTGGGGAGCAGCGAGACCAAGCTCTCTGACGATTCG 495
  || || || || || || || || || || || || || || || || || || || || || ||
Db 5757054 CGGTCTATCTCCGCGCGCAGCCACCGCGCGGCACCTCGCGCAGCGCCGAGGAGATCT 5757113
  || || || || || || || || || || || || || || || || || || || || || ||
QY 496 TGGAGGCGCTGTGGCGGCTGTGAGGAGGTGCCTGAAGCCTACGTCACATCCACACTGGCC 555
  || || || || || || || || || || || || || || || || || || || || || ||
Db 5757114 TGGGCGCGTGGCGAAGACCTTGGCGGTCCCGGAATCCAGCAGACGCGCTGCACCGCAC 5757173
  || || || || || || || || || || || || || || || || || || || || || ||
QY 556 TCAGTGGCAGTGGCGGTGGCGTTCGTGTGTGCAATCTCCGAGGCCCTGGCTGAAGGAGCGG 615
  || || || || || || || || || || || || || || || || || || || || || ||
Db 5757174 TCTCCGATCGGGCGCGCGGTACTTCTTCTATCTGTGAGGCCATGACGGACGCGGCA 5757233
  || || || || || || || || || || || || || || || || || || || || || ||
QY 616 TCAAGATGGCATGCCACAGCCTGGCCACCGCATCGCTGCCACGACCCCTGCTGGGGA 675
  || || || || || || || || || || || || || || || || || || || || || ||
Db 5757234 TCTGCTGGGAGTGGCCCGCGCAAGAGCGCAGCATCTGATCTCCAGTCCGCGATCGGCG 5757293
  || || || || || || || || || || || || || || || || || || || || || ||
QY 676 CGGCCAAGATGCTGTGTCACAGAGGCGCACACCCAGCCCGCTGCGCTCAGACGCTGCA 735
  || || || || || || || || || || || || || || || || || || || || || ||
Db 5757294 CCGGATCATGTCTCGGACAGCGGGAGCACCCCGTCAAGCTCCGAGAACGTGACGT 5757353
  || || || || || || || || || || || || || || || || || || || || || ||
QY 736 CCCCAGGTGGCACCAACATCTATGACTCCACGCCCTGGAGCAGGCGGGCTGCGAGCAG 795
  || || || || || || || || || || || || || || || || || || || || || ||
Db 5757354 CACCGCGCGCAGCAGCATCAACGCGATCCGCGAACTGGAGAACCAACGCGGTACGCGCG 5757413
  || || || || || || || || || || || || || || || || || || || || || ||
QY 796 CCACCATGAGCGCGCTGGAGGCTGCCACCTGCGCGGGCCAAAGGAGCT 841
  || || || || || || || || || || || || || || || || || || || || || ||
Db 5757414 CACTGATCGCGCGCTGGAGGAGCGCGCGCACCGCGCGGAACT 5757459
  || || || || || || || || || || || || || || || || || || || || || ||
```

RESULT 15

```
US-09-918-995-11309
; Sequence 11309, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11309
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(454)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-11309
```

```
Query Match          9.6%; Score 129; DB 12; Length 454;
Best Local Similarity 58.4%; Pred.No. 3.3e-22;
Matches 225; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 370 TCCCCCACAACACACGGGTGCTGGGGTCTTGCCCAACCTGCGCTGTGTGCTCCAGGAAG 429
  || || || || || || || || || || || || || || || || || || || || || ||
Db 69 TCCAGCCAGCCCCCAAGTGATTCGCTGATGACCAACACACCTGTGTGTGCGAGGAG 128
  || || || || || || || || || || || || || || || || || || || || || ||
QY 430 GGGCCATAGTATGGCGGGCGGCCACAGTGGGGAGCAGCGAGACCAAGCTCTCTGAGC 489
  || || || || || || || || || || || || || || || || || || || || || ||
Db 129 GCGCTACAGTGTACGCCACGGCACCCATGCCCTGGTGGAGGTGGCGAGCTCTCTGGAGC 188
  || || || || || || || || || || || || || || || || || || || || || ||
QY 490 ATCTGCTGAGGCGCTGTGAGGAGGTGCCTGAAGCCTACCTCCACATCCACA 549
  || || || || || || || || || || || || || || || || || || || || || ||
Db 189 AGCTCATGACAGCGTGGGCTTCTGCACTGAGGTGGAAGGAGGACCTCATCGATGCCGTCA 248
  || || || || || || || || || || || || || || || || || || || || || ||
QY 550 CTGGCGCTCAGTGGCAGTGGCGCTTCGTGTGTGCAATCTCCGAGGCGCTGGCTGAAG 609
  || || || || || || || || || || || || || || || || || || || || || ||
```

```
Db 249 CCGGGCTCAGTGGGAGCGGGGCTGCCCTATGCAATTCATGGCTCTGGAGCGCATTTGGCTGATG 308
QY 610 GAGCCGTCAAGATGGGCATGCCAGCAGCCTTGGCCCCACCGCATCGCTGCCCGCAGACCTGTC 669
  || || || || || || || || || || || || || || || || || || || || || ||
Db 309 GTGGGTGAAGATGGGTTTGGCACGGCGCCTTGGCAATCCAACTCGGGGCCCGAGCTTTC 368
  || || || || || || || || || || || || || || || || || || || || || ||
QY 670 TGGGAGCGGCCAAGATGCTGCTGACGAGGGCCAAACACCCAGCCAGCTGCGCTCAGACG 729
  || || || || || || || || || || || || || || || || || || || || || ||
Db 369 TGGAGGCTGCCAAGATGCTGCTGAGCTCGGAGCAGCATCCATGCCAGCTTAAAGGACAATG 428
  || || || || || || || || || || || || || || || || || || || || || ||
QY 730 TGTGCAACCCGGGTGGCACCACCAT 754
  || || || || || || || || || || || || || || || || || || || || || ||
Db 429 TCTGCTCCCTGGAGGAGCCACCAT 453
  || || || || || || || || || || || || || || || || || || || || || ||
```

Search completed: July 26, 2003, 07:28:27
Job time : 369 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2003, 04:52:43 ; Search time 2117 Seconds
(without alignments)
10235.976 Million cell updates/sec

Title: US-09-806-536A-29
Perfect score: 1338
Sequence: 1 ggtgagcgcagtctgtccga.....ggggcacgagcatatgggggt 1338

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_estl.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_fod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	884	66.1	1086	13	BM548059
2	871	65.1	906	9	AL526237
3	859.2	64.2	885	9	AL519338
4	832.2	62.2	1080	14	BM923064
5	786.4	58.8	873	13	BI837148
6	785.6	58.7	797	9	AL525111

7	763.8	57.1	773	13	BI827333	BI827333
8	753	56.3	856	13	BM014974	BM014974
9	737.6	55.1	794	9	AU141625	AU141625
10	701.2	52.4	1145	13	BM473687	BM473687
11	697.4	52.1	932	14	BQ36653	AGENCOURT
12	688.2	51.4	734	13	BI253559	BI253559
13	673.6	50.3	697	9	AL522085	AL522085
14	660	49.3	741	13	BM010281	BM010281
15	627.6	46.9	773	13	BI818555	BI818555
16	622.4	46.5	664	9	AI963473	AI963473
17	620.4	46.4	679	12	AG393676	AG393676
18	617	46.1	895	12	BF797243	BF797243
19	614.4	45.9	1036	13	BM545513	AGENCOURT
20	589.8	44.1	1024	13	BM465183	AGENCOURT
21	580.4	43.4	590	12	AG282890	AG282890
22	569.2	42.5	1271	11	AK002912	AK002912
23	567.8	42.4	823	13	BM019985	BM019985
24	557.4	41.7	595	10	BE21964	BE21964
25	552.8	41.3	1132	13	BM424026	BM424026
26	535.8	40.0	1265	11	AK013214	AK013214
27	535	40.0	803	14	BQ571032	BQ571032
28	532	39.8	744	10	BE267359	BE267359
29	532	39.8	902	14	BQ964436	BQ964436
30	527.6	39.4	788	13	BI254417	BI254417
31	525.6	39.3	874	13	BI415573	BI415573
32	520.8	38.9	1348	11	AK004291	AK004291
33	520.6	38.9	737	13	BM019007	BM019007
34	517.4	38.7	670	10	BE563597	BE563597
35	510.6	38.2	869	13	BI329060	BI329060
36	505.2	37.8	602	9	AL706316	AL706316
37	504.4	37.7	1178	13	BM461077	BM461077
38	495.6	37.0	722	14	BM945806	BM945806
39	486.6	36.4	563	9	AU129509	AU129509
40	482.6	36.1	752	14	BM945714	BM945714
41	465.4	34.8	736	13	BI903838	BI903838
42	457.8	34.2	476	9	AL556719	AL556719
43	453.6	33.9	660	14	BQ571218	BQ571218
44	449	33.6	741	13	BG967901	BG967901
45	430.8	32.2	619	12	BF787218	BF787218

ALIGNMENTS

RESULT 1
BM548059
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM548059
AGENCOURT_6531538
BM548059
BM548059.1
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1086)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12735 row: 1 column: 03
High quality sequence start: 13
High quality sequence stop: 685.

FEATURES

Location/Qualifiers

Db	120																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
----	-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Genoscope - Centre National de Sequencage	
BP 191 91006 EVRY cedex - France	
Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.	
Location/Qualifiers	
1. .885	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="CS0DB004Yf19"	
/clone_lib="LTI_NFL004_NBC2"	
/sex="male"	
/tissue_type="neuroblastoma cells"	
/lab_host="DH10B"	
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(OT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"	
BASE COUNT	147 a 290 c 156 t 2 others
ORIGIN	
Query Match	64.2%; Score 859.2; DB 9; Length 885;
Best Local Similarity	99.5%; Pred. No. 2.1e-183;
Matches	880; Conservative 2; Mismatches 0; Indels 2; Gaps 2;
QY	31 TGGCAGCTGCGGAGCCGCTCTCCGGCGCGTGGGCTGCTGGGCGCGGCGCATGGCGG 90
Db	1 TGGCAGCTGCGGA-CCGCTCTCCGGCGCGTGGGCTGCTGGGCGCGGCGCATGGCGG 59
QY	91 GGGCCATCGCCAGGGCTCATCAGCAGAGAAAGTGAAGCTCAGACATCTGGCCA 150
Db	60 GGGCCATCGCCAGGGCTCATCAGCAGAGAAAGTGAAGCTCAGACATCTGGCCA 119
QY	151 GTGCACCAACAGACAGGAACCTATGTCACTTTCAAGCTCTGGTTCGCGGACCGCACT 210
Db	120 GTGCACCAACAGACAGGAACCTATGTCACTTTCAAGCTCTGGTTCGCGGACCGCACT 179
QY	211 CCAACAGGAGGTGCTGCAGAGTGCCTGCTGCTCATCTTTGCCACCAAGCCTCATGTGC 270
Db	180 CCAACAGGAGGTGCTGCAGAGTGCCTGCTGCTCATCTTTGCCACCAAGCCTCATGTGC 239
QY	271 TGCCAGCTGCTCTGCAGAGTGGCTCCTGTGGTCAACACTGAACACATCTTGGTGTCCG 330
Db	240 TGCCAGCTGCTCTGCAGAGTGGCTCCTGTGGTCAACACTGAACACATCTTGGTGTCCG 299
QY	331 TGGCTGTGGGTGTCTCTGAGCACCTTGGAGGAGTCTGCTGCCCCCAACACAGGGGTGC 390
Db	300 TGGCTGTGGGTGTCTCTGAGCACCTTGGAGGAGTCTGCTGCCCCCAACACAGGGGTGC 359
QY	391 TGGGGTCTTGGCCAACTGCTGTGGTTCAGGAAGGGCCATAGTATGGCGGG 450
Db	360 TGGGGTCTTGGCCAACTGCTGTGGTTCAGGAAGGGCCATAGTATGGCGGG 419
QY	451 GCGGCGAGTGGGGAGCAGAGCAAGCTCTCTGACATCTGCTGGAGGCTGTGGGC 510
Db	420 GCGGCGAGTGGGGAGCAGAGCAAGCTCTCTGACATCTGCTGGAGGCTGTGGGC 479
QY	511 GGTGTGAGGAGTGGCTTGAAGCCTACGTGACATCCACACTGGGCTCAGTGGCAGTGGCG 570
Db	480 GGTGTGAGGAGTGGCTTGAAGCCTACGTGACATCCACACTGGGCTCAGTGGCAGTGGCG 539
QY	571 TGGCCTTGTGTGCTATCTCCGAGGCGCTGGCTGAAGGAGCCCTCAAGATGGCATGC 630
Db	540 TGGCCTTGTGTGCTATCTCCGAGGCGCTGGCTGAAGGAGCCCTCAAGATGGCATGC 599
QY	631 CCAGCAGCTGGCCACCGCATCTGCTGCCAGACCTCTGCTGGGAGCGCCCAAGATGCTGC 690
Db	600 CCAGCAGCTGGCCACCGCATCTGCTGCCAGACCTCTGCTGGGAGCGCCCAAGATGCTGC 659


```
QY 498 GAGCGCTGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCTC 557
|||||
Db 487 GAGCGCTGTGGCGGTGTGAGGAGTGCCTGAAGCCTACGTGCACATCCACACTGGCTC 546
|||||
QY 558 AGTGGCAGTGGCGTGGCTTGTGTGTGTCATTCCTCCGAGGCCCTGGCTGAAGGAGCGGTC 617
|||||
Db 547 AGTGGCAGTGGCGTGGCTTGTGTGTGTCATTCCTCCGAGGCCCTGGCTGAAGGAGCGGTC 606
|||||
QY 618 AAGATGGGATGCCAGAGCCTGGCCACCGCATCCGCTGCCAGACCCCTGTGGGAGC 677
|||||
Db 607 AAGATGGGATGCCAGAGCCTGGCCACCGCATCCGCTGCCAGACCCCTGTGGGAGC 666
|||||
QY 678 GCCAAGATGCTGCTGCAGAGGGCCAAACACCCAGCTGGCTCAGAGCTGTGCACC 737
|||||
Db 667 GCCAAGATGCTGCTGCAGAGGGCCAAACCCAGCTGGCTCAGAGCTGTGCACC 726
|||||
QY 738 CCGGTGGCACCACCATCTATGGACTCCAGCCCTGGAGGAGCGG 784
|||||
Db 727 CCGGTGGCACCACCATCTATGGACTCCAGCCCTGGAGGAGCGG 773
|||||

RESULT 8
BM014974
LOCUS 603640923F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5417260 5',
DEFINITION mRNA sequence.
ACCESSION BM014974.1 GI:16529328
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 856)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12064 row: m column: 05
High quality sequence stop: 783.
Location/Qualifiers
1. 856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5417260"
/cloned_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="organ: breast; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 158 a 267 c 290 g 141 t
ORIGIN

Query Match 56.3%; Score 753; DB 13; Length 856;
Best Local Similarity 97.9%; Pred. No. 1.9e-159;
Matches 837; Conservative 0; Mismatches 10; Indels 8; Gaps 7;

QY 18 CGAGGCAACAAGATGCGAGCTCGGAGCGGCTCCGCGCGCGTGGCTTGTGGGCGG 77
|||||
Db 2 CGAGGCAACAAGATGCGAGCTCGGAGCGGCTCCGCGCGCGTGGCTTGTGGGCGG 61
|||||
QY 78 GGCCCGCATGGCGGGGCCATCGCGCAGGCGCTCATCAGAGCAGGAAAGTGAAGCTCAG 137
|||||
```

```
Db 62 GGCCGATGGCGGGGCCATCGGCA -GCCCTCATCAGAGCAGGAAAAGTGAAGCTCAG 120
|||||
QY 138 CACATACTGGCAGTGCACCAACAGACAGAGAACCTATGTCACTTTCAAGCTCTGGTTC 197
|||||
Db 121 CACATACTGGCAGTGCACCAACAGACAGAGAACCTATGTCACTTTCAAGCTCTGGTTC 180
|||||
QY 198 CGSACCACACACTCCAAACAGGAGGTGCTGCAGAGTGCCTGCTCATCTTTTGCAC 257
|||||
Db 181 CGSACCACACACTCCAAACAGGAGGTGCTGCAGAGTGCCTGCTCATCTTTTGCAC 240
|||||
QY 258 AAGCCTCATGTGCTGCAGCTGTCTTGGCAGAGTGGCTCTGTGTGCTACCACTGAACAC 317
|||||
Db 241 AAGCCTCATGTGCTGCAGCTGTCTTGGCAGAGTGGCTCTGTGTGCTACCACTGAACAC 300
|||||
QY 318 ATCTTGGTGTCCGTGGCTGTCTGTGAGCACCCTGGAGAGTGTCTGCCCCCA 377
|||||
Db 301 ATCTTGGTGTCCGTGGCTGTCTGTGAGCACCCTGGAGAGTGTCTGCCCCCA 360
|||||
QY 378 AACACACGGGTGCTGGGGTCTTGGCCAAACCTGCCCTGTGTGCTCCAGGAAGGGCCATA 437
|||||
Db 361 AACACACGGGTGCTGGGGTCTTGGCCAAACCTGCCCTGTGTGCTCCAGGAAGGGCCATA 420
|||||
QY 438 GTGATGGCGGGGCCGCCACGTGGGAGAGCAGACCAAGCTCTCTGAGCATCTGTGTG 497
|||||
Db 421 GTGATGGCGGGGCCGCCACGTGGGAGAGTGGCTGAAGCCTACGTGCACATCCACACTGGCCTC 480
|||||
QY 498 GAGGCGCTGTGGCGGTGTGAGGAGTGGCTGAAGCCTACGTGCACATCCACACTGGCCTC 557
|||||
Db 481 GAGGCGCTGTGGCGGTGTGAGGAGTGGCTGAAGCCTACGTGCACATCCACACTGGCCTC 540
|||||
QY 558 AGTGGCAGTGGCTGGCTTGTGTGTGCAATTCCTCGAGGCCCTGGCTGAAGGAGCGGTC 617
|||||
Db 541 AGTGGCAGTGGCTGGCTTGTGTGTGCAATTCCTCGAGGCCCTGGCTGAAGGAGCGGTC 600
|||||
QY 618 AAGATGGGATGCCAGCAGCCTTGCCCCACCGCATCGCTGCCAGACCCCTGTCTGGGAGC 677
|||||
Db 601 AAGATGGGATGCCAGCAGCCTTGCCCCACCGCATCGCTGCCAGACCCCTGTCTGGGAGC 659
|||||
QY 678 GCCAAGATGCTGTGCAGAGGGCCACACCCAGCCAGCTCGCTC -AGACCTGTGCA - 735
|||||
Db 660 GCAAAGATGCTGTGCAGAGGGCCACACCCAGGCCAGCTCGCTCAAGCATGTGTGCAC 719
|||||
QY 736 CCCGGGTGGCACCACCATCTAT -GGACTCCAGCCCTGGAGCAGGCG - -GGCTGTGCGAG 792
|||||
Db 720 CCCGGGTGGCACCACCATCTATGGGACTCCAGCCCTGGAGCAGGCGCGGCTTGGAG 779
|||||
QY 793 CAGCCACCATGAGCGCGTGGAGGCTGCCACCTGCCGGGCCAAGGAGCTCAGCAGAAAGT 852
|||||
Db 780 CAGCAACCATGAGCGCGGGGAGGCTGCAACCTGCCCGG -CAAGAAGGTGAGCAGAAAGT 838
|||||
QY 853 AGGCTGGGCTCTGGC 867
|||||
Db 839 AGGCTGGGCTCTGGC 853
|||||

RESULT 9
BM014625
LOCUS THYR01 Homo sapiens cDNA clone THYR01000934 5', mRNA
DEFINITION sequence.
ACCESSION AU141625
VERSION AU141625
KEYWORDS GI:11003146
SOURCE EST.
ORGANISM Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 794)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
```

**JOURNAL
COMMENT**

Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project: 5' - & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and

FEATURES	SOURCE
1. High Accuracy: The model achieves a high accuracy rate, consistently performing well across various datasets and tasks.	1. High Accuracy: The model achieves a high accuracy rate, consistently performing well across various datasets and tasks.
2. Scalability: The model is designed to handle large-scale data and complex tasks, making it suitable for enterprise-level applications.	2. Scalability: The model is designed to handle large-scale data and complex tasks, making it suitable for enterprise-level applications.
3. Interpretability: The model's decisions are transparent and explainable, allowing users to understand the underlying logic and reasoning.	3. Interpretability: The model's decisions are transparent and explainable, allowing users to understand the underlying logic and reasoning.
4. Robustness: The model is highly resilient to adversarial attacks and data poisoning, ensuring reliable performance in real-world scenarios.	4. Robustness: The model is highly resilient to adversarial attacks and data poisoning, ensuring reliable performance in real-world scenarios.
5. Efficiency: The model is optimized for fast inference and low resource consumption, making it ideal for deployment on edge devices.	5. Efficiency: The model is optimized for fast inference and low resource consumption, making it ideal for deployment on edge devices.
6. Flexibility: The model can be easily adapted to different domains and tasks, providing a versatile solution for various use cases.	6. Flexibility: The model can be easily adapted to different domains and tasks, providing a versatile solution for various use cases.
7. Security: The model incorporates advanced security measures to protect sensitive data and prevent unauthorized access.	7. Security: The model incorporates advanced security measures to protect sensitive data and prevent unauthorized access.
8. Integration: The model seamlessly integrates with existing systems and workflows, enabling a smooth transition to AI-powered solutions.	8. Integration: The model seamlessly integrates with existing systems and workflows, enabling a smooth transition to AI-powered solutions.
9. Customization: The model offers a high degree of customization, allowing users to tailor the model to their specific requirements.	9. Customization: The model offers a high degree of customization, allowing users to tailor the model to their specific requirements.
10. Support: The model is backed by a dedicated support team, ensuring timely assistance and troubleshooting for any issues.	10. Support: The model is backed by a dedicated support team, ensuring timely assistance and troubleshooting for any issues.

```

1. .794
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="THYROID00934"
/clone_lib="ThyR1"
/tissue_type="thyroid gland"
/note="Vector: pME185FL3"
142 a 248 c 264 g 133 t 7 others

```

Query Match 55.1%; Score 737.6; DB 9; Length 794;
Best Local Similarity 97.7%; Pred. NO. 5.5e-156;
Matches 765; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

QY	15	GTCCGAGGCAACAAGATGGCAGCTGGGAGCCGCTCTCCGCGGCGGTGGGCTTCGTGGGC	74
DB	12	GTNCGAGGCAACAAGATGGCAGCTGGGAGCCGCTNTCCGCGGCGGTGGGCTTCGTGGGC	71
QY	75	GGGGGCCCATGCGGGGGCCATCGCGCAGGGCCTCATCAGACGAGGAAAGTGGAACT	134
DB	72	GGGGGCCCATGCGGGGGCCATCGCGCAGGGCCTCATCAGACGAGGAAAGTGGAACT	131
QY	135	CAGCACATCTGGCCAGTCACCAACAGACAGGAACCTATGTCACTTTCAAGCTCTGGGT	194
DB	132	CAGCACATCTGGCCAGTCACCAACAGACAGGAACCTATGTCACTTTCAAGCTCTGGGT	191
QY	195	TGCGGGACCAAGCACTCCACACGAGGAGTGCTCAGAGCTGCTGCTGTCATCTTTGGCC	254
DB	192	TGCGGGACCAAGCACTCCACACGAGGAGTGCTCAGAGCTGCTGCTGTCATCTTTGGCC	251
QY	255	ACCAAGCTCATGTGCTGCCAGCTGCTCGGACAGAGTGGCTCCTGCTGTCACACACTGAA	314
DB	252	ACCAAGCTCATGTGCTGCCAGCTGCTCGGACAGAGTGGCTCCTGCTGTCACACACTGAA	311
QY	315	CACATCTTTGGTCCGTGGCTGGTGGGGTGTTCTTGAGCACCTCGGAGAGCTGCTGCC	374
DB	312	CACATCTTTGGTCCGTGGCTGGTGGGGTGTTCTTGAGCACCTCGGAGAGCTGCTGCC	371
QY	375	CCAACACAGGGTGCTCGGGGTCTTTGCCCAACCTGCCCTGTGTGGTCAGAGAGGGCC	434
DB	372	CCAACACAGGGTGCTCGGGGTCTTTGCCCAACCTGCCCTGTGTGGTCAGAGAGGGCC	431
QY	435	ATAGTGATGGCGGGGCCCCACGTGGGGAGCAGGAGACCAAGCTCC-TGCAGCACTC	493
DB	432	ATAGTGATGGCGGGGCCCCACGTGGGGAGCAGGAGACCAAGCTCCTTGCAGCACTC	491
QY	494	GCTGGAGCCCTGTGGGCGGTGTGAGGAGTGGCTGAAAGCCTACGTCGACATCCACACTGG	553
DB	492	GCTGGAGCCCTGTGGGCGGTGTGAGGAGTGGCTGAAAGCCTACGTCGACATCCACACTGG	551
QY	554	CCTCAGTGGCAGTGGCGTGGCCCTTCGTGTGTGTCATCTCCGAGGCCCTGGCTGAAGAGC	613
DB	552	CCTCAGTGGCAGTGGCGTGGCCCTTCGTGTGTGTCATCTCCGAGGCCCTGGCTGAAGAGC	611
QY	614	CGTCAAGATGGGATGCCACGACGCTGGCCACCGCATTCGCTGCCACAGACCCCTGCTGGG	673
DB	612	CGTCAAGATGGGATGCCACGACGCTGGCCACCGCATTCGCTGCCACAGACCCCTGCTGGG	671
QY	674	GACGGCCAGATGCTGTGCAGAGGGCCAAACCCAGCCGCTGTGGCTCAGACGCTGTG	733

Db	672	GACGGCCAAAGATGCTGCTGCACGAGGGCCAAACACCCAGCCAGCTNGCTCAGACGTGTG	731
Qy	734	CA-CCCCGGGTGGCACCACCACCATCTATGGACTCCACGCCCTGGAGCAGGGCGGGCTCGGAG	792
Db	732	CACCCCGGGTGGGACCAACCATCTATGGACTCCACGCCCTGGACAGGGCCGGCTG	791
Qy	793	CAG 795	
Db	792	CAG 794	
RESULT 10			
BM473687			
LOCUS	BM473687	1145 bp	mRNA
DEFINITION	AGENCOURT_6466593 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562066		linear EST 05-FEB-2002
ACCESSION	5', mRNA sequence.		
VERSION	BM473687		
KEYWORDS	BM473687.1 GI:18522729		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
COMMENT	1 (bases 1 to 1145)		
	NIH-MGC http://mgc.nci.nih.gov/		
	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: c9apbs-femail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LIAM12290 row: n column: 19		
	High quality sequence stop: 503.		
FEATURES	Location/Qualifiers		
source	1..1145		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5562066"		
	/clone_lib="NIH_MGC_88"		
	/tissue_type="duodenal adenocarcinoma, cell line"		
	/note="Organ: small intestine; Vector: pCMV-SPORT6;		
	Site.1: NotI; Site.2: SalI; Cloned unidirectionally;		
	oligo-dT primed. Average insert size 1.767 kb. library		
	enriched for full-length clones and constructed by Life		
	Technologies. Note: this is a NIH_MGC Library."		
BASE COUNT	194 a 392 c 356 g 195 t		8 others
ORIGIN			
Query Match	52.4%	Score 701.2;	DB 13; Length 1145;
Best Local Similarity	96.5%;	Pred. No. 1..le-147;	
Matches 727;	Conservative	0; Mismatches 24;	Indels 2; Gaps 1;
Qy	24	AACAAGATGGCAGCTCGCGAGCGCTTCGCGCGCGCGTTCGTTGGCGCGCGCGCGC	83
Db	1	AACAAGATGGCAGCTCGCGAGCGCTTCGCGCGCGCGTTCGTTGGCGCGCGCGCGC	60
Qy	84	ATGGCGGGGGCCATCGCGCAGGGCCCTATCATCAGACGAGGAAAAGTGGAGCTCAGCACATA	143
Db	61	ATGGCGGGGGCCATCGCGCAGGGCCCTATCATCAGACGAGGAAAAGTGGAGCTCAGCACATA	120
Qy	144	CTGGCCAGTGCACCAACACACAGGAACCTATGTCACTTTCAGGCTCTGGTTGCCGGACC	203
Db	121	CTGGCCAGTGCACCAACACACAGGAACCTATGTCACTTTCAGGCTCTGGTTGCCGGACC	180
Qy	204	ACGCACTCCAACAGGAGGTGCTGCAGAGCTGCCTGCCTCATCTTTGCCACCAAGCCT	263
Db	181	ACGCACTCCAACAGGAGGTGCTGCAGAGCTGCCTGCCTCATCTTTGCCACCAAGCCT	240

QY 264 CATGTGCTGCCAGCTCTCTCGCAGAGTGGCTCTGTGGTCACCACTGAACACATCTTG 323
|||||
Db 241 CATGTGCTGCCAGCTCTCTCGCAGAGTGGCTCTGTGGTCACCACTGAACACATCTTG 300
|||||
QY 324 GTGTCCGTGGCTGCTGGGTGTCTCTGAGCACCTGGAGGAGCTGTGCCCCCAACACA 383
|||||
Db 301 GTGTCCGTGGCTGCTGGGTGTCTCTGAGCACCTGGAGGAGCTGTGCCCCCAACACA 360
|||||
QY 384 CGGGTCTGGGGTCTTTGCCCAACCTGCTGTGTGGTCCAGAAAGGGGCCATAGTGATG 443
|||||
Db 361 CGGGTCTGGGGTCTTTGCCCAACCTGCTGTGTGGTCCAGAAAGGGGCCATAGTGATG 420
|||||
QY 444 GCSCGGGGCGCCACCTGGGGAGCAGCAGACCAAGCTCTCGACATCTGCTGGAGGCC 503
|||||
Db 421 GCSCGGGGCGCCACCTGGGGAGCAGCAGACCAAGCTCTCGACATCTGCTGGAGGCC 480
|||||
QY 504 TGTGGCGGTGTGAGGAGTGGCTGAAGCTTACGTGACATCCACACTGGCCTCAAGATG 523
|||||
Db 541 AGTGGCGGTGGCTTCTGATTTCTCGAGGCCCTGGCTGAAGGAGCCGCTCAAGATG 600
|||||
QY 624 GGCATGCCAGCAGCTGGCCACCGCATCGCTGGCCAGACCTGCTGGGGAGCGGCAAG 683
|||||
Db 601 GGCATGCCAGCAGCTGGCCACCGCATCGCTGGCCAGACCTGCTGGGGAGCGGCAAG 660
|||||
QY 684 ATGCTGCTGACAGAGGGGCAACACCCAGCCAGCTGCGCTGAGACGTGTGCACCCC - -GG 741
|||||
Db 661 ATGCTGCTGACAGAGGGGCAACACCCAGCCAGCTGCGCTGAGACGTGTGCACCCC - -GG 720
|||||
QY 742 GTGGCACCACCATCTATGACTCCAGCCCTGG 774
|||||
Db 721 GGGGACCCCATCTATGACTCCAGCCCTGG 753
|||||

RESULT 11

BQ436653 932 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7801851 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6050678
5', mRNA sequence.
ACCESSION BQ436653
VERSION BQ436653.1 GI:21175729
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 932)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

AUTHORS

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DRP

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13303 row: e column: 15

High quality sequence stop: 565.

FEATURES

source

Location/Qualifiers
1..932
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6050678"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"

RESULT 12

B1253559

LOCUS

DEFINITION

602973485F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112692 5',

mRNA sequence.

ACCESSION

B1253559

VERSION

B1253559.1

KEYWORDS

EST.

SOURCE

human.

/note="organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 163 a 311 c 306 g 152 t
ORIGIN

Query Match 52.1%; Score 697.4; DB 14; Length 932;
Best Local Similarity 97.0%; Pred. No. 7e-147;
Matches 732; Conservative 0; Mismatches 21; Indels 2; Gaps 2;
QY 41 GGAGCGCTCTCCGGGCGCTGGGTCTGTGGGGCGGGCGGCATGGCGGGGCGCATCGC 100
|||||
Db 1 GGAGCGCTCTCCGGGCGCTGGGTCTGTGGGGCGGGCGGCATGGCGGGGCGCATCGC 60
|||||
QY 101 GGAGGCGCTCATCAGAGCAGGAGAAAGTGAAGCTCAGCACATACTGGCCAGTGCACCAAC 160
|||||
Db 61 GGAGGCGCTCATCAGAGCAGGAGAAAGTGAAGCTCAGCACATACTGGCCAGTGCACCAAC 120
|||||
QY 161 AGACAGGAACCTATGTCACTTTCAAGCTCTGGGTGGCGGACACGACACTCCAAACCAAGA 220
|||||
Db 121 AGACAGGAACCTATGTCACTTTCAAGCTCTGGGTGGCGGACACGACACTCCAAACCAAGA 180
|||||
QY 221 GGTGCTGCAGAGCTGCTGCTGCTCATCTTTGCCAACCAAGCCTCATGTGCTGCCAGCTGT 280
|||||
Db 181 GGTGCTGCAGAGCTGCTGCTGCTCATCTTTGCCAACCAAGCCTCATGTGCTGCCAGCTGT 240
|||||
QY 281 CCTGCAGAGTGGTCTCTGTGGTCACCACTGAACACATCTTGGTGTCCGTGGCTGGCTGG 340
|||||
Db 241 CCTGCAGAGTGGTCTCTGTGGTCACCACTGAACACATCTTGGTGTCCGTGGCTGGCTGG 300
|||||
QY 341 GGTGCTCTGAGCACCTGGAGAGCTGCTGCCCAACACACACGGGTGCTGCCGGTCTT 400
|||||
Db 301 GATGCTCTGAGCACCTGGAGAGCTGCTGCCCAACACACACGGGTGCTGCCGGTCTT 360
|||||
QY 401 GCCCAACCTGCCCTGTGTGGTCCAGAAAGGGGCCATAGTATGGCGGGGCGGCCACGCT 460
|||||
Db 361 GCCCAACCTGCCCTGTGTGGTCCAGAAAGGGGCCATAGTATGGCGGGGCGGCCACGCT 420
|||||
QY 461 GGGGAGCAGCAGACCAAGCTCTCGCAGCATCTGCTGGAGGCGCTGTGGGCGGTGTGAGGA 520
|||||
Db 421 GGGGAGCAGCAGACCAAGCTCTCGCAGCATCTGCTGGAGGCGCTGTGGGCGGTGTGAGGA 480
|||||
QY 521 GGTGCTGAAGCCTACGCTCGACATCCACACTGGCCTCAGTGGCAGTGGCGCTTGGT 580
|||||
Db 481 GGTGCTGAAGCCTACGCTCGACATCCACACTGGCCTCAGTGGCAGTGGCGCTTGGT 540
|||||
QY 581 GTGTGCATTTCCGAGGCGCTGGCTGAAGAGGCGCTCAAGATGGGCATGCCACAGCGCT 640
|||||
Db 541 GTGTGCATTTCCGAGGCGCTGGCTGAAGAGGCGCTCAAGATGGGCATGCCACAGCGCT 600
|||||
QY 641 GGGCCACCGCATCGCTGCCAGACCTGCT-GGGGACGGCCCAAGATGCTGTGCACGAGG 699
|||||
Db 601 GGGCCACCGCATCGCTGCCAGACCTGCTGTGGGGAGCCGCGCAAGATGCTGTGCACGAGG 660
|||||
QY 700 GCCAACACCCAGCCAGCTGGCCTCAGACGTGTGCACCCC-GGGTGGGACCACTCAT 758
|||||
Db 661 GCCAACACCCAGCCAGCTGGCCTCAGACGTGTGCACCCCAGGGGGGACCACTCAT 720
|||||
QY 759 GGACTCCAGCGCTTGAGAGCGGCGGTGGGAGC 793
|||||
Db 721 GGACTCCCGCGCTGGGAAACCAAGGGGCGGCTGC 755
|||||

B1253559 734 bp mRNA linear EST 17-JUL-2001

602973485F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112692 5',

mRNA sequence.

B1253559 734 bp mRNA linear EST 17-JUL-2001

602973485F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112692 5',

mRNA sequence.

B1253559 734 bp mRNA linear EST 17-JUL-2001

602973485F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112692 5',

mRNA sequence.

B1253559 734 bp mRNA linear EST 17-JUL-2001

602973485F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112692 5',

mRNA sequence.

ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D.
COMMENT	Email: cgabbs@remail.nih.gov cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Incyte Genomics, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM11273 row: j column: 21 High quality sequence stop: 734.
FEATURES	Location/Qualifiers
source	1. .734 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5112692" /clone_lib="NIH_MGC_12" /tissue_type="cervical carcinoma cell line" /lab_host="DH10B" /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.4 kb. Library prepared by Life Technologies."
BASE COUNT	127 a 228 c 253 g 126 t
ORIGIN	
Query Match	51.4%; Score 688.2; DB 13; Length 734;
Best Local Similarity	98.6%; Pred. No. 7.4e-145;
Matches	715; Conservative 0; Mismatches 8; Indels 2; Gaps 3
Qy	36 GCTGCGGAGCGCTCTCCGGCGCGCTGGGCTTCGTGGCGGGCGGCATGGCGGGGGCC 95
Db	
Qy	11 GGTGCGGAGCGCTCTCCGGCGCGCTGGGCTTCGTGGCGGGCGGCATGGCGGGGGCC 70
Db	
Qy	96 ATCGCGCAGGGCTCATCAGACGAGGAAAGTGAAGCTCAGCACATACCTGGCCAGTGCA 155
Db	
Qy	71 ATCGCGCAGGGCTCATCAGACGAGGAAAGTGAAGCTCAGCACATACCTGGCCAGTGCA 130
Db	
Qy	156 CCAACAGACAGGAACCTATGTCATCTTCAAGCTCTGGGTTGCCGACACGACATCCAA 215
Db	
Qy	131 CCAACAGACAGGAACCTATGTCATCTTCAAGCTCTGGGTTGCCGACACGAGGCGCA 190
Db	
Qy	216 CAGGAGGTGCTCAGAGCTGCTGCTGTCATCTTTGCCACCAAGCTCATGTGCTGCCA 275
Db	
Qy	191 CAGGAGGTGCTCAGAGCTGCTGCTGTCATCTTTGCCACCAAGCTCATGTGCTGCCA 250
Db	
Qy	276 GCTGTCTGGCAGAGTGGCTCCTGTGCTCACCACCTGAACACATCTTGGTGTCCGTGGCT 335
Db	
Qy	251 GCTGTCTGGCAGAGTGGCTCCTGTGTCACCACTGAACACATCTTGGTGTCCGTGGCT 310
Db	
Qy	336 GCTGGGTGTCCTGAGCACCCCTGGAGAGCTGCTGCCCCCAACACAGGGTGCTCGCG 395
Db	
Qy	311 GCTGGGTGTCCTGAGCACCCCTGGAGAGCTGCTGCCCCCAACACAGGGTGCTCGCG 370
Db	
Qy	396 GTCTTGTCCCAACCTCCCTGTGTGTCAGGAAGGGCCATAGTATGGCGGGGGCGCC 455
Db	
Qy	371 GTCTTGTCCCAACCTCCCTGTGTGTCAGGAAGGGCCATAGTATGGCGGGGGCGCC 430
Db	
Qy	456 CACGTGGGAGCAGCAGAACCAAGCTCCTGCAGCATCTGCTGGAGGCCTGTGGCGGTGT 515
Db	
Qy	431 CACGTGGGAGCAGCAGAACCAAGCTCCTGCAGCATCTGCTGGAGGCCTGTGGCGGTGT 490
Db	
Qy	516 GAGGAGTGCCTGAAGCCTAGCTCGACATCCACACTGGCCTCACTGGCAGTGGCGTGGCC 575
Db	
Qy	491 GAGGAGTGCCTGAAGCCTAGCTCGACATCCACACTGGCCTCACTGGCAGTGGCGTGGCC 550
Db	

```
Db 120 ACTGGCCAGTGCACCAACAGACAGAACCTATGTCTACCTTCAAGCTCTGGGTTGCCGAC 179
QY 203 CAGGCACCTCAACAGGAGGTGCTGCAGAGCTGCCTGCTGTCATCTTTGGCCACCAAGCC 262
Db 180 CAGGCACCTCAACAGGAGGTGCTGCAGAGCTGCCTGCTGTCATCTTTGGCCACCAAGCC 239
QY 263 TCATGTGCTGCCAGTGTCTCTGGCAGAGGTGCTCTGTGGTCCACCACTGAACACATCTT 322
Db 240 TCATGTGCTGCCAGTGTCTCTGGCAGAGGTGCTCTGTGGTCCACCACTGAACACATCTT 299
QY 323 GGTGTCCTGGTGTCTGGGTTCTCTGTAGCACCCTGGAGGAGTGTCTGCCCCCAACAC 382
Db 300 GGTGTCCTGGTGTCTGGGTTCTCTGTAGCACCCTGGAGGAGTGTCTGCCCCCAACAC 359
QY 383 ACGGTGCTGCGGGTCTTCCCAACCTGCCCTGTGTGTCTCCAGGAGGGGCTATAGTAT 442
Db 360 ACGGTGCTGCGGGTCTTCCCAACCTGCCCTGTGTGTGTCTCCAGGAGGGGCTATAGTAT 419
QY 443 GCGCGGGGCG - GCGCCACGTGGGAGCAGCGAGACCAAGCTCTCTGAGCATCTGCTGGAGG 501
Db 420 GCGCGGGGCGCGCCACGTGGGAGCAGCGAGACCAAGCTCTCTGAGCATCTGCTGGAGG 479
QY 502 CTTGTGGGCGGTGTGAGGAGTGTCTGAAGCCCTAGTGCATATCCACACTGCGCTCAGTG 561
Db 480 CTTGTGGGCGGTGTGAGGAGTGTCTGAAGCCCTAGTGCATATCCACACTGCGCTCAGTG 539
QY 562 GCAGTGGCTGCGCTTCGTGTGTGATCTCTCCGAGGCCCTGGCTGAAGAGCGGCTCAAGA 621
Db 540 GCAGTGGCTGCGCTTCGTGTGTGATCTCTCCGAGGCCCTGGCTGAAGAGCGGCTCAAGA 599
QY 622 TGGGCATGCCAGCAGCCCTGGCCACCGCATCGCTGCCAGACCCTGCTGGGAGCGCCA 681
Db 600 TGGGCATGCCAGCAGCCCTGGCCACCGCATCGCTGCCAGACCCTGCTGGGAGCGCCA 659
QY 682 AGATGCTGTGCACGAGGCGCCACACCCAGCCAGCTG 719
Db 660 AGATGCTGTGCACGAGGCGCCACACCCAGCCAGCTG 697
```

RESULT 14

```
BM010281 741 bp mRNA linear EST 30-OCT-2001
LOCUS 603631154F1 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:544893 5',
DEFINITION mRNA sequence.
```

```
ACCESSION BM010281
VERSION BM010281.1 GI:16524635
KEYWORDS EST.
```

SOURCE

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE

```
1 (bases 1 to 741)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
```

AUTHORS

```
TITLE Contact: Robert Strausberg, Ph.D.
JOURNAL Email: cgabbs-remail.nih.gov
COMMENT Tissue Procurement: DCTD/Drp
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1924 row: 1 column: 14
High quality sequence stop: 676.
```

FEATURES

source

```
1..741
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:544893"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
```

RESULT 15

```
BI818555
LOCUS 603033078F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:517478 5',
DEFINITION
```

```
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 133 a 226 c 255 g 127 t
ORIGIN
```

```
Query Match 49.3%; Score 660; DB 13; Length 741;
Best Local Similarity 96.2%; Pred. No. 1.7e-138;
Matches 708; Conservative 0; Mismatches 25; Indels 3; Gaps 3;
QY 15 GTCCGAGCAACAAGATGGCAGTGCAGGAGCCGTCTCCGGGCGCGTTCGTGGGC 74
Db 7 GTCCGAGCAACAAGATGGCAGTGCAGGAGCCGTCTCCGGGCGCGTTCGTGGGC 66
QY 75 GCGGGCCGATGCGGGGGCCATCGCGAGGGCCCTCATCAGAGCAGGAAAGTGAAGCT 134
Db 67 GCGGGCCGATGCGGGGGCCATCGCGAGGGCCCTCATCAGAGCAGGAAAGTGAAGCT 126
QY 135 CAGCACATATCTGGCCAGTGCACCAACAGAGAACCTATGTCACCTTTCAAGCTCTGGGT 194
Db 127 CAGCACATATCTGGCCAGTGCACCAACAGAGAACCTATGTCACCTTTCAAGCTCTGGGT 186
QY 195 TGCGGAGCAGCGCACTCCAAACAGAGAGTGTCTGAGAGTGCCTGCTGCTATCTTTGCC 254
Db 187 TGCGGAGCAGCGCACTCCAAACAGAGAGTGTCTGAGAGTGCCTGCTGCTATCTTTGCC 246
QY 255 ACCAAGCCTCATGTGCTGCCAGTGTCTGCGAGAGTGTCTGCTGAGAGTGTCTGCTGCC 314
Db 247 ACCAAGCCTCATGTGCTGCCAGTGTCTGCGAGAGTGTCTGCTGAGAGTGTCTGCTGCC 306
QY 315 CACATCTTTGGTGTCTGCTGCCAGTGTCTGCGAGAGTGTCTGCTGAGAGTGTCTGCTGCC 374
Db 307 CACATCTTTGGTGTCTGCTGCCAGTGTCTGCGAGAGTGTCTGCTGAGAGTGTCTGCTGCC 366
QY 375 CCAACACACAGGCTGTGCGGGTCTTGCCCAACCTGCGCTGTGTGTCAGGAGGGGCC 434
Db 367 CCAACACACAGGCTGTGCGGGTCTTGCCCAACCTGCGCTGTGTGTCAGGAGGGGCC 426
QY 435 ATAGTGTGCGCGGGGCGCCACGTGGGAGCAGCAGACCAAGCTCTCTCAGCATCTG 494
Db 427 ATAGTGTGCGCGGGGCGCCACGTGGGAGCAGCAGATGACCAACTCTCTCAGCATCTG 486
QY 495 CTGAGGCGCTGTGGGCGGTGTGAGAGTGCCTGAAGCCTACGTCGACATCCACACTGGC 554
Db 487 CTGAGGCGCTGTGGGCGGTGTGAGAGTGCCTGAAGCCTACGTCGACATCCA-ACTGGC 545
QY 555 CTCAGTGCAGTGGCGTGGCCCTTCTGCTGTCATCTCCGAGGCGCTGGCTGAA-GGAGC 613
Db 546 CTCAGTGCAGTGGCGTGGCCCTTCTGCTGTCATCTCCGAGGCGCTGGCTGAAAGCAGC 605
QY 614 CGTCAAGATGGCATGCCACAGCCTGGCCACCGCATCTGCTCCACAGACCTCTGCTGG 673
Db 606 CGTCAAGATGGCATGCCACAGCCTGGCCACCGCATCTGCTCCACAGACCTCTGCTGG 665
QY 674 GACGCCAAGATGCTGTGTCAGAGGGGCAACACCCA-GCCAGCTGCGCTCAGACCTGT 732
Db 666 GACGCCAAGATGCTGTGTCAGAGGGGCAACACCCAGGCGCAAGCTGCGCTCAGACCTGT 725
QY 733 GCACCCCGGTGGCAC 748
Db 726 GCACCCCGGTGGCAC 741
```

mRNA sequence.
BI818555
VERSION BI818555.1 GI:15929661
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 773)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11433 row: 1 column: 19
High quality sequence stop: 763.
Location/Qualifiers
1..773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5174178"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."

FEATURES

Source

BASE COUNT 146 a 237 c 257 g 133 t
ORIGIN
Query Match 46.9%; Score 627.6; DB 13; Length 773;
Best Local Similarity 96.2%; Pred.No. 3.6e-131;
Matches 729; Conservative 0; Mismatches 19; Indels 10; Gaps 8;
QY 15 GTCCGAGGCAACAGATGGCAGCTCGGAGCGCTCTCCGGCGCGCTGGGCTTCGTGGGC 74
DB 12 GTCCGAGGCAACAGATGGCAGCTCGGAGCGCTCTCCGGCGCGCTGGGCTTCGTGGGC 71
QY 75 GCGGGCCGCATGGCGGGGCCATCGCGCAGGGCCCTCATCAGAGCAGGAAAGTGGAAAGCT 134
DB 72 GCGGGCCGCATGGCGGGGCCATCGCGCA-GGCCCTCATCAGAGCAGGAAAGTGGAAAGCT 130
QY 135 CAGCACATCTGGCCAGTGCACACAGACAGAGGACCTATGTCACCTTCAAGCTCTGGGT 194
DB 131 CAGCACATCTGGCCAGTGCACACAGAGGAGGTGCTGCAGAGCTGCCCTGCTCACTTTGGGT 190
QY 195 TGCCGGACACGACCTCAACACAGGAGGTGCTGCAGAGCTGCCCTGCTCGTCATCTTTGGC 254
DB 191 TGCCGGACACGACCTCAACACAGGAGGTGCTGCAGAGCTGCCCTGCTCACTTTGGC 250
QY 255 ACCAAGCCTCATGTGCTGCCAGTGTCTTGGCAGAGGTGGCTCCTGTGGTCAACACTGAA 314
DB 251 ACCAAGCCTCATGTGCTGCCAGTGTCTTGGCAGAGGTGGCTCCTGTGGTCAACACTGAA 310
QY 315 CACATCTTGGTGTCCGTGGCTGTGGGTGCTCTGAGCAGGAGGAGTGGTGGTGGTGGTGGT 374
DB 311 CACATCTTGGTGTCCGTGGCTGTGGGTGCTCTGAGCAGGAGGAGTGGTGGTGGTGGTGGT 370
QY 375 CCAACACACGGGTGCTCGGGGTCTTGCCAACTGCGCTGTGTGGTCCAGGAGGGGCC 434
|||||

Db 371 CCAACACACGGGTGCTCGGGGTCTTGCCAACTGCGCTGTGTGGTCCAGGAGGGGC 430
QY 435 ATAGTGATGGCGGGGCCCGCCACAGTGGGGAGGAGACCAAGCTCTCTGCAGCA--TC 492
|||||
Db 431 ATAGTGATGGCGGGGCCG-CACGTGGGAGGAGGAGACCAAGCTCTCTGCAGCATCTC 489
493 TGCTGGAGGCTGTGGCGGTGTGAGGAGTGGCTTGAAGCTTACGTGCGACATCCACACTG 552
490 TGCTGGAGGCTGTGGCGGTGTGAGGAGTGGCTTGAAGCTTACGTGCGACATCCACACTG 549
553 GCCTCAGTGGCAG-TGGCGTGGCTTTCGTGTGTCATTCTCCAGGCGCTGGCTGAAGGA 611
550 GCCTCAGTGGCAGTGGCGTGGCTTTCGTGTGTCATTCTCCAGTGGCTGAAGGA 609
612 GCGTCAAGATGGCATGCCACAGCCTGG-CCCACCGCATGCTGCCAGACCTGCT 670
610 GCGTCAAGATGGCATGCCACAGCCTGGAGTGGAGTGGCTGCCAGAGCTGCT 669
671 GGGGA-CGGCAAGATGCTGCTGCAC--GAGGGCCCAACACCCAGCCAGCTGCGCTCAG- 726
670 GGGGACCGCAAGATGCTGCTGCACAGAGGGGCCAAGAACCCAGGCCAAGCTGCGCTCAGA 729
QY 727 ACCTGTGCACCCCGGGTGGCACCACCATCTATGGACTC 764
730 ACCTGTGCACCCCGGGTGGCACCACCATCTATGGACTC 767
|||||

Search completed: July 26, 2003, 07:20:42
Job time : 2127 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: July 21, 2003, 09:48:21 ; Search time 80 Seconds
(without alignments)
705.711 Million cell updates/sec

Title: US-09-806-536A-14
Perfect score: 1386
Sequence: 1 MAAAEPSRRVFGVAGRMA.....AATMSAVEAATCRAKELSRK 274

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phase:*
10: sp.plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1386	100.0	274	4	Q9H896
2	1381	99.6	274	4	Q96HX4
3	1163	83.9	274	11	Q9DCC4
4	1160	83.7	274	11	Q8R0P9
5	921	66.5	248	11	Q9D0X2
6	613	44.2	279	5	Q21544
7	577.5	41.7	320	4	Q96C36
8	571.5	41.2	320	11	Q922Q4
9	560.5	40.4	273	5	Q9VEJ3
10	560.5	40.4	309	11	Q922W5
11	533.5	38.5	270	17	Q9HH99
12	505.5	36.5	280	5	Q9V3F8
13	503	36.3	274	10	Q9AYM3
14	493.5	35.6	280	5	Q96B43
15	480	34.6	270	16	Q97E64
16	477	34.4	319	4	Q9Y5J4

ALIGNMENTS

RESULT 1

Q9H896 PRELIMINARY; PRT; 274 AA.
ID Q9H896
AC Q9H896; 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THYR01000934 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYROID;
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT *NEO human cDNA sequencing project.*;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK023914; BAB14721.1;
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1.
DR PROSITE: PS00521; P5CR; 1.
SQ SEQUENCE 274 AA; 28663 MW; 846FDEC603F3B548 CRC64;

Query Match 100.0%; Score 1386; DB 4; Length 274;

Best Local Similarity 100.0%; Pred. No. 1.4e-100; Mismatches 0; Indels 0; Gaps 0;

Matches 274; Conservative 0;

QY 1 MAAAEPSRRVFGVAGRMAIAOGLIRAGKVEAQHILASAPTDRNLCHFQALGCRTH 60

|||||

Db 1 MAAAEPSRRVFGVAGRMAIAOGLIRAGKVEAQHILASAPTDRNLCHFQALGCRTH 60

|||||

QY 61 SNOEVLQSCLLVIFATKPHVLPVLAIEAVPVVTTTHILVSVAAAGVSLTLELLPPNTRV 120

|||||

Db 61 SNOEVLQSCLLVIFATKPHVLPVLAIEAVPVVTTTHILVSVAAAGVSLTLELLPPNTRV 120

|||||

QY 121 LRVLPNLCVQVQGAIVMARGHVSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGS 180
Db |||||||
QY 121 LRVLPNLCVQVQGAIVMARGHVSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGS 180
Db |||||||
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db |||||||
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db |||||||
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db |||||||
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db |||||||

RESULT 2

Q96HX4 ID Q96HX4 PRELIMINARY; PRT; 274 AA.
AC Q96HX4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 28.6 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007993; AAH07993.1; -
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1.
DR PROSITE: PS00521; P5CR; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 274 AA; 28649 MW; 846FD9B60183B048 CRC64;

Query Match 99.6%; Score 1381; DB 4; Length 274;
Best Local Similarity 99.6%; Pred. No. 3.5e-100;
Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MAAEPSPRRVGVGAGRMAGAIAGLIRAGKVEAQHILASAPTRDNLCHFQALGCRTH 60
Db |||||||
QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEPVVTTEHILVSVAAAGVSLSTLEELLPPNTRY 120
Db |||||||
QY 121 LRVLPNLCVQVQGAIVMARGHVSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGS 180
Db |||||||
QY 121 LRVLPNLCVQVQGAIVMARGHVSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGS 180
Db |||||||
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db |||||||
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db |||||||
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db |||||||
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db |||||||

RESULT 3

Q9DCC4 ID Q9DCC4 PRELIMINARY; PRT; 274 AA.
AC Q9DCC4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 1110058B13rik protein.
GN 1110058B13rik.
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bernaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK002912; BAB22451.1; -
DR MGD; MGI:1913444; 1110058B13rik.
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1.
SQ SEQUENCE 274 AA; 28694 MW; FE3892C01C6068A5 CRC64;

Query Match 83.9%; Score 1163; DB 11; Length 274;
Best Local Similarity 82.5%; Pred. No. 3.7e-83;
Matches 226; Conservative 23; Mismatches 25; Indels 0; Gaps 0;
QY 1 MAAEPSPRRVGVGAGRMAGAIAGLIRAGKVEAQHILASAPTRDNLCHFQALGCRTH 60
Db |||||||
QY 1 MAATMSEPRRVGVGAGRMAGAIAGLIRAGKVEAQHILASAPTRDNLCHFQALGCRTH 60
Db |||||||
QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEPVVTTEHILVSVAAAGVSLSTLEELLPPNTRY 120
Db |||||||
QY 61 SNEHVLSQCLVIFATKPHVLPVLAEPVVTTEHILVSVAAAGVSLSTLEELLPPNTRY 120
Db |||||||
QY 121 LRVLPNLCVQVQGAIVMARGHVSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGS 180
Db |||||||
QY 121 LRVSPNLCVQVQGAIVMARGHVSSETKLLQHLLEACGRCCEVPEAYVDIHTGLSGS 180
Db |||||||
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db |||||||
QY 181 VAFVCAFSEALAEAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGHPAQLRSDVCTPGGT 240
Db |||||||
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db |||||||
QY 241 TIYGLHALBOGGLRAATMSAVEAATCRAKELSRK 274
Db |||||||

RESULT 4

Q8R0P9 ID Q8R0P9 PRELIMINARY; PRT; 274 AA.
AC Q8R0P9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RIKEN cDNA 1110058B13 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;

```
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC026536; AAB26536.1; -. E01FB7133B45BD7C CRC64;
SQ SEQUENCE 274 AA; 28721 MW; 28721 MW; 28721 MW; 28721 MW;

Query Match 83.7%; Score 1160; DB 11; Length 274;
Best Local Similarity 82.1%; Pred. No. 6.3e-83;
Matches 225; Conservative 24; Mismatches 25; Indels 0; Gaps 0;

QY 1 MAEAEPSPRRVGVGAGRMAGAIAOGLIRAGKVEAQHILASAPTDRLNLCFQALGCRTH 60
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MAATMSPRRVGVGAGRMAGAEIARGLIQAGKVEAQHILASAPTDRLNLCFQALGCRTH 60
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 SNOEVLQSCLLVIFATKPHVLPVLAEVAPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 SNEVLQSCLLVIFATKPHVLPVLAEVAPVVTTEHILVSAAGVSLSTLEELLPPNTRV 120
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 LRVLPNLCVVOEGAIVMARGHVSGSETKLLQHLLEACRCCEEYVYVDIHTGLSGSG 180
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 LRVLPNLCVVOEGAIVMARGHVSGSETKLLQHLLEACRCCEEYVYVDIHTGLSGSG 180
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 181 VAFVCAFESEALAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 VAFVCTFESEALAGAIKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 240
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 241 TIYGLHAL 248
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 237 TCWNHSM 244
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
Q21544 PRELIMINARY; PRT; 279 AA.
AC Q21544;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE M153.1 protein.
GN M153.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews P.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
   investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z67995; CAA91943.1; -.
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1
DR TIGRFAMs: TIGR00112; proc; 1.
SQ SEQUENCE 279 AA; 29165 MW; C8CADA4BC4EB20E6 CRC64;

Query Match 44.2%; Score 613; DB 5; Length 279;
Best Local Similarity 46.1%; Pred. No. 3.3e-40;
Matches 123; Conservative 56; Mismatches 86; Indels 2; Gaps 1;

QY 10 RVGVGAGRMAGAIAOGLIRAGKVEAQHILASAP--TDRLNLCFQALGCRTHSNOEVLQ 67
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 2 KIGFIGAGKMAQALAGLINSGRITADNIASSPKRDEVLQDCKALGLNTHDIAEVVQ 61
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY 68 SCLLVIFATKPHVLPVLAEVAPVVTTEHILVSAAGVSLSTLEELLPPNTRVLPN 127
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 62 KSDVFLAVKPHVSVKVAISEIAPLSKEHLVSVAGITIRIESLLPTKRVKRWMPNT 121
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY 128 PCVVOEGAIVMARGHVSGSETKLLQHLLEACRCCEEYVYVDIHTGLSGGVAFVCAF 187
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 122 PSVVRAGASAFAMGSCNRDGDATVEKLLSTVGFAVEPEIHDVPTGLSGSPMYFAV 181
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY 188 SEALAGAVKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGT 247
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

Db 182 IEGLADGVKVGVLPRDLALKLAAYTLGAAKMVLEGTIHPAQLKDDVQSPAGSSVYGMHK 241
QY 248 LEQGLRAATMSAVEAATCRAKELSRK 274
DB 242 LESGGLKGYLMDAVEAATNRSRATGDK 268

RESULT 7

Q96C36 ID Q96C36 PRELIMINARY; PRT; 320 AA.
AC Q96C36;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to pyroline 5-carboxylate reductase isoform (Hypothetical
DE 33.6 kDa protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=CERVIX;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014868; AAH14868.1; -;
DR EMBL; BC020553; AAH20553.1; -;
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 320 AA; 33637 MW; 9348455A6ACCE722 CRC64;

Query Match 41.7%; Score 577.5; DB 4; Length 320;
Best Local Similarity 46.4%; Pred. No. 2.2e-37;
Matches 124; Conservative 48; Mismatches 88; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGATAGLIRAGKVEAQHILASAPTDRNLCHFQAL---GCRTHSNOEVLQ 67
DB 3 VGFAGQAYALARGFTAAAGLSAHKIIASSP-EMNLPTVSALRRMGVNLTRSNKTYK 61
QY 68 SCLLVIFATKPHVLPVLAEAPVVTTEHILYSVAAGVSLTLELL---PNTNRLVRL 124
DB 62 HSDVLFVAVKPHIIPILDEIGADVQARHIVVSCAAGVTISSVEKKLMFAFPAPKVICM 121
QY 125 PNLPCVQBGAIWMARGHRVGSSETKLQHLLEACRCCEVPEAYVDIHTGLSGGVAFV 184
DB 122 TNPVVOBGATVYATGTHALVEDGQLLEQMLSSVGFCTEVEDLIDAVTGLSGGPAYA 181
QY 185 CAFSEALARGAVKMGMPSSLAHRIAATLLGTAKMLHGGHQAOLRSVCPGGTTIYG 244
DB 182 FMALEDALADGVKMGVPRRLALQGAQLLGAAKMLLDSHQPCQLKDNVCSPPGGATIHA 241
QY 245 LHALEQGGGLRAATMSAVEAATCRAKEL 271
DB 242 LHFLSGGGFRSLINAVEASCIRTREL 268

RESULT 8

Q922Q4 ID Q922Q4 PRELIMINARY; PRT; 320 AA.
AC Q922Q4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to pyroline 5-carboxylate reductase isoform.
GN 1810018M05RIK.
OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006882; AAH06882.1; -;
DR MGD; MGI:1916301; 1810018M05RIK.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; UNKNOWN_1.
SQ SEQUENCE 320 AA; 33659 MW; 64DD7F96FB8C4BC5 CRC64;

Query Match 41.2%; Score 571.5; DB 11; Length 320;
Best Local Similarity 46.1%; Pred. No. 6.6e-37;
Matches 123; Conservative 48; Mismatches 89; Indels 7; Gaps 3;

QY 11 VGFVAGRMAGATAGLIRAGKVEAQHILASAPTDRNLCHFQAL---GCRTHSNOEVLQ 67
DB 3 VGFAGQALACALARGFTAAAGLSAHKIIASSP-DMDLPTVSALRRMGVNLTRSNKDTVR 61
QY 68 SCLLVIFATKPHVLPVLAEAPVVTTEHILYSVAAGVSLTLELL---PNTNRLVRL 124
DB 62 HSDVLFVAVKPHIIPILDEIGADVQARHIVVSCAAGVTISSVEKKLMFAFPAPKVICM 121
QY 125 PNLPCVQBGAIWMARGHRVGSSETKLQHLLEACRCCEVPEAYVDIHTGLSGGVAFV 184
DB 122 TNPVVOBGATVYATGTHALVEDGQLLEQMLSSVGFCTEVEDLIDAVTGLSGGPAYA 181
QY 185 CAFSEALARGAVKMGMPSSLAHRIAATLLGTAKMLHGGHQAOLRSVCPGGTTIYG 244
DB 182 FMALEDALADGVKMGVPRRLAVRLGAQLLGAAKMLLDSHDHPGQLKDNVCSPPGGATIHA 241
QY 245 LHALEQGGGLRAATMSAVEAATCRAKEL 271
DB 242 LHFLSGGGFRSLINAVEASCIRTREL 268

RESULT 9

Q9VEJ3 ID Q9VEJ3 PRELIMINARY; PRT; 273 AA.
AC Q9VEJ3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG5840 protein (re62767p).
GN CG5840
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaese R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo K., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., M.G.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan M., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Aqbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Patel A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patela S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003718; AAP55428.1; -.
DR EMBL; AY071558; AAL49180.1; -.
DR FlyBase; FBgn0038516; CG5840.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR TIGRFAMs; TIGR00112; proC; 1.
DR PROSITE; PS00521; P5CR; 1.
SQ SEQUENCE 273 AA; 28099 MW; 563754C62ACE77AB CRC64;

Query Match 40.4%; Score 560.5; DB 5; Length 273;
Best Local Similarity 44.2%; Pred. No. 4e-36;
Matches 118; Conservative 50; Mismatches 96; Indels 3; Gaps 3;

QY 10 RVGVGAGRWAGATAOGLIRAGKVEAQHILASA-PTDR-NLCHFOALGCRTHSHNOEVLQ 67
DB 6 KIGFLGGNNWAKALAKGLAAGLAKPNTLIASVHPADKLSQFSGLGVETVTKNAPVWQ 65
QY 68 SCLVIFATPKPHVLPVLAEPVVTTEHILSVAGVSLSTLEELLPPNTRVLRVLPNL 127
DB 66 QSDVVFVSVPQVPSVLSLQIP-LSSGKFLSVAMGITLSTIESSLSPOAIRVMPNL 124
QY 128 PCVVQEGAIYVARGRHVGSSETKLLQHLLEACRCCEVPAYVDIHTGLSGSGVAFVCF 187
DB 125 PAVVSCGCSFVFRGSKATDADITQKLSQVGTCEPVDSEQLDVTALSGSPAYVFM 184
QY 188 SEALAEAGVKGMPSSLAHRIAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGLHA 247
DB 185 IEALDAGVHMGPRDRLAYRLASQTVLGAHGMVDSGMHPGLKDGVTSPAGSTAALRQ 244
QY 248 LEQGLRAATMSAVEAATCRAKELSRK 274
DB 245 LELSGFRAAVSGAVEQATLCRCQISGK 271

RESULT 10
Q922W5
ID Q922W5 PRELIMINARY; PRT; 309 AA.
AC Q922W5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Similar to pyrroline-5-carboxylate reductase 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006727; AAH06727.1; -.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR TIGRFAMs; TIGR00112; proC; 1.
DR PROSITE; PS00521; P5CR; UNKNOWN 1.
SQ SEQUENCE 309 AA; 32373 MW; A3CD24AACDD53DEF CRC64;

Query Match 40.4%; Score 560.5; DB 11; Length 309;
Best Local Similarity 43.2%; Pred. No. 4.6e-36;
Matches 115; Conservative 54; Mismatches 92; Indels 5; Gaps 2;

QY 11 VGVGAGRWAGATAOGLIRAGKVEAQHILASAP--TDRNLCHFOALGCRTHSHNOEVLQS 68
DB 3 VGFIGAGQDAFALAGFTTAAGVLAHKIMASSPDDQATVSALRKIGVNLTPHNKETVRH 62
QY 69 CLVIFATPKPHVLPVLAEPVVTTEHILSVAGVSLSTLEELL---PPNTRVLRVLP 125
DB 63 SDVFLAVRPHIIPFILDIEGANIEDRHVVSCAAGVTINSIEKKLTAFQAPKPVIRWT 122
QY 126 NLPVVQEGAIYVARGRHVGSSETKLLQHLLEACRCCEVPAYVDIHTGLSGSGVAFVCF 185
DB 123 NTPVVVREGVTYATGTTHAQVEDGRLEQMLGMSVGFCTEVEDLIDAVTGLSGSPAYAF 182
QY 186 AFSEALAEAGVKGMPSSLAHRIAQTLLGTAKMLLHEGHPAQLRSDVCTPGGTTIYGL 245
DB 183 TALDALADGGVKGMLPRRLVRLGALLGAARKMLLDSEQHPSQLKDNVCSPPGATIHAL 242
QY 246 HALEQGLRAATMSAVEAATCRAKEL 271
DB 243 HVLESGGFRSLINAVEASCIRTREL 268

RESULT 11
Q9HH99
ID Q9HH99 PRELIMINARY; PRT; 270 AA.
AC Q9HH99;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ProC (Pyrroline-5-carboxylate reductase).
DE PROC OR MA4102.
GN Methanosarcina acetivorans.
OS Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A;
RA Zhang J.K., White A.K., Kuettner H.C., Boccazzi P., Metcalf W.W.;
RT "Molecular and genetic analysis of proline biosynthesis in the
methanogenic archaeon *Methanosarcina acetivorans* C2A.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown R.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame J.A.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,

Db 184 VEMADAANVINGWREKAYKFAAQSVLGAAKMILETGEHPGKLDKDDVCSPSGTTIEAVYA 243
Qy 248 LEQGLRAATMSAVEAATCRAKELSRK 274
Db 244 LEKSGFRASVIAAVDACIKKSKLMSSK 270

Search completed: July 21, 2003, 09:55:13
Job time : 83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 21, 2003, 09:43:01 ; Search time 23 Seconds
(without alignments)
494.109 Million cell updates/sec

Title: US-09-806-536A-14

Perfect score: 1386

Sequence: 1 MAAAEPSRRVGVGAGRMA.....AATMSAVEATCRAKELSRK 274

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	584.5	42.2	319	1	PROC_HUMAN
2	502.5	36.3	273	1	PROC_PEA
3	498	35.9	278	1	PROC_ACTCH
4	492	35.5	274	1	PROC_SOYBN
5	475	34.3	276	1	PROC_ARATH
6	423.5	30.6	269	1	PROC_ECOLI
7	412.5	29.8	267	1	PROC_SYNY3
8	390.5	28.2	263	1	PROC_TREPA
9	387	27.9	261	1	PROC_TRETH
10	383.5	27.7	294	1	PROC_MYCLE
11	378	27.3	297	1	PROH_BACSU
12	366	26.4	295	1	PROC_MYCTU
13	362.5	26.2	270	1	PROC_CORGL
14	362	26.1	299	1	PROC_CAEL
15	345.5	24.9	272	1	PROC_PSEAE
16	344	24.8	282	1	PROC_SCHPO
17	337.5	24.4	278	1	PROI_BACSU
18	321	23.2	286	1	PROC_YEAST
19	318.5	23.0	285	1	PROC_AQUAE
20	310.5	22.4	311	1	PROC_NEUCR
21	307	22.2	278	1	PROC_VIBAL
22	305.5	22.0	320	1	PROC_ZALAR
23	301.5	21.8	271	1	PROC_HAEIN
24	198	14.3	272	1	PROC_BACSU
25	189.5	13.7	273	1	CME4_BACSU
26	164	11.8	257	1	PROC_HELPU
27	159	11.5	251	1	PROC_HELPY
28	125.5	9.1	257	1	PROC_METSM
29	114	8.2	294	1	GARR_ECOLI
30	99.5	7.2	388	1	ALR_MYCAY
31	97	7.0	1503	1	MRP6_HUMAN
32	96	6.9	401	1	CSD_PSEAE
33	96	6.9	528	1	SERA_MYCTU

34 95 6.9 4393 1 PGBM_HUMAN
35 94.5 6.8 388 1 ALR_MYCLE
36 93.5 6.7 349 1 GPDA_MYCLE
37 93 6.7 528 1 SERA_MYCLE
38 92.5 6.7 2436 1 ABC2_HUMAN
39 92 6.7 278 1 HBD_DEIRA
40 92 6.6 386 1 ALR_MYCTU
41 92 6.6 400 1 NTRB_AZOB
42 92 6.6 482 1 MURC_PASMU
43 90 6.5 334 1 GPDA_MYCTU
44 90 6.5 388 1 ALR3_RHILO
45 89.5 6.5 2301 1 POLG_TMEVD

ALIGNMENTS

RESULT 1

PROC_HUMAN
ID PROC_HUMAN STANDARD; PRT; 319 AA.
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PYCR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112821; PubMed=1730675;
RT "Cloning human pyroline-5-carboxylate reductase cDNA by
complementation in Saccharomyces cerevisiae.";
RL J. Biol. Chem. 267:871-875(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = L-pyrroline-5-
carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBUNIT: HOMODECAMER OR HOMODECAMER.
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; M7836; AAA36407.1; -
CC EMBL; BC001504; AAH01504.1; -
CC PIR; A41770; A41770.
CC Genew; HGNC:9721; PYCR1.
CC MIM; 179035; -
CC InterPro; IPR000304; P5CR.
CC Pfam; PF01089; P5CR; 1.
CC TIGRFAMs; TIGR00112; proC; 1.
CC PROSITE; PS00521; P5CR; 1.
KW Oxidoreductase; Proline biosynthesis; NADP.
FT CONFLICT 155 155 T -> S (IN REF. 2).
SQ SEQUENCE 319 AA; 33374 MW; F5E74B5BDFB475EF CRC64;

Query Match 42.2%; Score 584.5; DB 1; Length 319;
Best Local Similarity 46.4%; Pred. No. 2.9e+38;
Matches 124; Conservative 47; Mismatches 89; Indels 7; Gaps 3;


```
Db 117 D-WAGHSRIRVMPNTPAAYGEAASVMSLGATGEGELITLFLGAIGKIWADEKLEF 175
QY 172 IHTGLSGGVAFYCAFSALAEAGVAKMGMPSSLAHRIAQTLLGTAKMLLHGHQHPAQLR 231
Db 176 AVTGLSGGPAYIFLAEALADGVAGLPRDLASLQTVLGAASWVARGKHGPKQLK 235
QY 232 SDVCTPGGTTTYGLHALEOGLRAATMSAVEAATCRAKELSRK 274
Db 236 DDVASAGGTTIAGHLEKEGFRGTLMLNAVVSATKRSQEIFKR 278

RESULT 4
PROC_SOYBN
ID PROC_SOYBN STANDARD; PRT; 274 AA.
AC P17817;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90340278; PubMed=2199815;
RA Delauney A.J., Verna D.P.S.;
RT "A soybean gene encoding delta 1-pyrroline-5-carboxylate reductase
RT was isolated by functional complementation in Escherichia coli and is
RT found to be osmoregulated.";
RL Mol. Gen. Genet. 221:299-305(1990).
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = 1-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL PLANT TISSUES, BUT MOSTLY IN
CC NODULES.
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16352; CAA34401.1; -.
DR PIR; S10186; S10186.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR TIGRFAMs; TIGR00112; PROC; 1.
DR PROSITE; PS00521; P5CR; 1.
KW Oxidoreductase; Proline biosynthesis; NADP.
SQ SEQUENCE 274 AA; 28586 MW; 933CFCD7598B63 CRC64;

Query Match 35.58; Score 492; DB 1; Length 274;
Best Local Similarity 39.68; Pred. No. 3.4e-31;
Matches 107; Conservative 51; Mismatches 96; Indels 16; Gaps 3;

Qy 11 VGFVAGRMAGATAOGLIRAGKVEAQAHLASAPTDRLNCH-----FOALGCRTHSN 62
Db 13 LGFIGACKMAESIARGAVSGVLPPSR-----RTAVHFNLRARGAFESFGVTVLPSN 65
Qy 63 QEVQLSCLLVIFATKPHVLAIEVAPVVTTEHILVSVAAVGSLSLTLELLPPNTRVLR 122
Db 66 DDVVRSDVVVLSVKPOLKDVVSKLTPLTKHLVSVAAAGTKLKLQEQE-WAGNDRFIR 124
Qy 123 VLPLNLCVQEGALVAVRGHRVGSSETKLLQHLLEACGRCEEVPEYVVDIHTGLSGGVA 182
Db 125 VMPTNPAVGQAAVMSLGGSATGEDGNITIAQLFGSIGKTIWKAEEKYFDAITGLSGGPA 184
```

```
QY 183 FYCAFSEALAEAGVAKMGMPSSLAHRIAQTLLGTAKMLLHGHQHPAQLRSDVCTPGGTTI 242
Db 185 VYVLAIEALADGVAGLPRDLASLQTVLGAASWVARGKHGPKQLKDDVTSPEGTTI 244
QY 243 YGLHALEOGLRAATMSAVEAATCRAKELS 272
Db 245 TGIHELENGFRGTLMLNAVVAALKRSRELS 274

RESULT 5
PROC_ARATH
ID PROC_ARATH STANDARD; PRT; 276 AA.
AC P54904;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PROC1 OR AT5G14800 OR T9L3.100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94294559; PubMed=8022935;
RA Verbruggen N., Villarroel R., van Montagu M.;
RT "Osmoregulation of a pyroline-5-carboxylate reductase gene in
RT Arabidopsis thaliana.";
RL Plant Physiol. 103:771-781(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RA Verbruggen N., Villarroel R., Hua X., van Montagu M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dente M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volktaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
RA Latham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,
RA Rampersperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staveren M., Dirkse W., Mooijman P., Klein Lankhorst R.,
RA Weitznegger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Giehl J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:823-826(2000).
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = 1-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
```



```

RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugiura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -|- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = L-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -|- PATHWAY: Proline biosynthesis; third (last) step.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90916; BAA18679.1; -
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1.
DR TIGRfam: TIGR00112; proC; 1.
DR PROSITE; PS00521; P5CR; 1.
DR OXIDOREDUCTASE; Proline biosynthesis; NADP; Complete proteome.
SQ SEQUENCE 267 AA; 27883 MW; FD9C428047F391B0 CRC64;

Query Match 29.8%; Score 412.5; DB 1; Length 267;
Best Local Similarity 36.2%; Pred. No. 4.6e-25;
Matches 98; Conservative 50; Mismatches 110; Indels 13; Gaps 4;

QY 10 RVGFGAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNL-CHFOALGCRTHSNQEVLSQS 68
DB 4 QLGIIIGGVMAEAILARLIAETKYEPIIIVGEPHGCARDYQVYVSPDNOEAVN 63
QY 69 CLLVIFATPKPHVLPVAVLEA- - - - -PVVTTTHILVSVAAVSLSTLELLPNTRVLRV 123
DB 64 SEVLLAVRPQVLDRLVSLAGSAGNRPV- - - - -ISILAGVSLQRIQKGFDPDA-IIRA 116
QY 124 LPLPCVQVQEGATVMAARGHVGSGSETKLLQHLLEAGRCCEVEPYAVVDIHTGLSGGVAF 183
DB 117 MPTATVAGMTATIAANKMVEPDQLAKAIFSAVGNVVEPNLMDAVTGVSGSGPAY 176
QY 184 VCAFSALAEAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGTTIY 243
DB 177 VALMIEALADGGVLAGLPRAIAQKLAQVLTGTAELIKETEEHPAQIKDKVTSFGTTIA 236
QY 244 GHALSQGGGLRAATMSAVAAATCRAKELSRK 274
DB 237 GVAVLEKMGFRSAIEAVRAAYRRSOELGKK 267

RESULT 8
PROC_TREPA STANDARD; PRT; 263 AA.
AC P27771; O83775;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PROC OP TP0797.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;

```

```

RX MEDLINE=90264287; PubMed=2188947;
RA Gherardini F.C., Hobbs M.M., Stamm L.V., Bassford P.J. Jr.;
RT "Complementation of an Escherichia coli proc mutation by a gene cloned
RT from Treponema pallidum.";
RL J. Bacteriol. 172:2996-3002(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -|- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = L-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -|- PATHWAY: Proline biosynthesis; third (last) step.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -|- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73825; AAA27478.1; -
DR EMBL; AE001250; AAC65760.1; -
DR PIR; S27734; S27734.
DR TIGR; TP0797; -
DR InterPro: IPR000304; P5CR.
DR Pfam: PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; 1.
DR OXIDOREDUCTASE; Proline biosynthesis; NADP; Complete proteome.
FT CONFLICT 1 65
FT FT 77 93
FT CONFLICT VLRDRQSFQGVKLVLSLA -> GTARSPFQESAISC (IN
FT REF. 1).
FT CONFLICT VRAAL -> CRWLS (IN REF. 1).
FT SEQUENCE 263 AA; 27645 MW; ED4AD7C54BAF9D61 CRC64;

Query Match 28.2%; Score 390.5; DB 1; Length 263;
Best Local Similarity 38.1%; Pred. No. 2.3e-23;
Matches 99; Conservative 41; Mismatches 11; Indels 9; Gaps 5;

QY 11 VGFVGAGRMAGAIAGLIRAGKVEAQHILASAPTDRLNL-CHFOALGCRTHSNQEVLSQS 69
DB 3 VGFLGFGAGRMAGLAEGLVHAGALQAQVYACALNQEKRAOCTSLGIGACASVQELVQKS 62
QY 70 LLVIFATPKPHVLPVAVLEA- - - - -PVVTTTHILVSVAAVSLSTLELLPN- - - - -TRVLRLVP 125
DB 63 EWIFLAVKPSQISTVLRDRQSF- - - - -QGVKLVLSLAAGMSCAAYEALFAADPHQGIHRLSLP 120
QY 126 NLPVQVQEGATVMAARGH-VGSSETKLLQHLLEAGRCCEVEPYAVVDIHTGLSGGVAFV 184
DB 121 NLPQVARG-VIIAEARHTLHDEHAALLAVLRTVAQVVDVTFAYFIAGYIACAPAF 179
QY 185 CAFSEALAEAGVAKMGMPSSLAHRIAQAOTLLGTAKMLLHEGQHPAQLRSDVCTPGGTTIY 244
DB 180 AQFEALADAGVRYGLARDQAYRLAAHMLEGTAALIOHSGVHPAQLKDRVCSAGSTIRG 239
QY 245 LHALEQGGGLRAATMSAVEAA 264

```

```
Db 240 VLAEEQGLRAVIAVRAA 259
RESULT 9
PROC_THETH
ID PROC_THETH STANDARD; PRT; 261 AA.
AC P34893;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PROC.
OS Thermus thermophilus.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Thermales;
OC Thermaceae; Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB27;
RX MEDLINE=94168609; PubMed=8123043;
RA Hoshino T., Kosuge T., Hidaka Y., Tabata K., Nakahara T.;
RT "Molecular cloning and sequence analysis of the proc gene encoding
RT delta 1-pyrroline-5-carboxylate reductase from an extremely
RT thermophilic eubacterium Thermus thermophilus";
RL Biochem. Biophys. Res. Commun. 199:410-417(1994).
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = 1-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D25413; BAA05001.1; -
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; 1.
DR OXidoreductase; Proline biosynthesis; NADP.
SQ SEQUENCE 261 AA; 27819 MW; 3435C3A8F4299B3E CRC64;
Query Match 27.9%; Score 387; DB 1; Length 261;
Best Local Similarity 37.7%; Pred. No. 4.2e-23;
Matches 101; Conservative 47; Mismatches 102; Indels 18; Gaps 9;
QY 10 RVGFVAGRMAGAIAGLIRAG--KVEAQHILASAPTDRLNCHFGALGCR-TTHSNQEV 66
Db 2 RLAIVGLKMGKRSILKALGERFLRPEVGVLRTP-ERSRELAEPFGVRPLRADLGM 60
QY 67 QSCLLVIFATKPHVLPVLAIEVAPVVTTEHIL--VSVAGVSLSTLEELLPNTRVL 124
Db 61 ER---VLIIVQPPDFPALAPEIA-----HRLGYISIMAGISTVLAARL-DNRRVRAM 111
QY 125 PNLPCVVQEGAIVMARGHVGSSE-TKLQHLLEACGRCCEVPEYVDIHTGLSGGVAF 183
Db 112 PNLAVVIGESSALTALKAREADLAFARLFAIVGVGYEPEHFLDAFTGMSASAPY 171
QY 184 VCAFSBALAEGAVKMGMPSSLAHRTAAOTLLGTAKMLHEGQHPAQLRSDVCTPGGTTIY 243
Db 172 LAVVAEALADAGVKMGMPRALRLAALDAALATGELL--KGRHPAQVKDEVASPGGTTIH 229
QY 244 GLHLEQGLRAATNSAVEAATCRAKEL 271
Db 230 GLHALEARAVRAAFYEAVEAATRRGHEL 257
RESULT 10
PROC_MYCLE
ID PROC_MYCLE STANDARD; PRT; 294 AA.
AC P46725;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase (EC 1.5.1.2) (P5CR) (P5C reductase).
GN PROC OR ML2430 OR B2168_C2_211.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBDJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) = 1-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U00018; AAA17233.1; -
DR EMBL; AL583925; CAC31947.1; -
DR Leproma; ML2430; -
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR PROSITE; PS00521; P5CR; 1.
DR OXidoreductase; Proline biosynthesis; NADP; Complete proteome.
SQ SEQUENCE 294 AA; 30237 MW; EA8606C9CBBB6D9D CRC64;
Query Match 27.7%; Score 383.5; DB 1; Length 294;
Best Local Similarity 32.8%; Pred. No. 8.9e-23;
Matches 94; Conservative 57; Mismatches 113; Indels 23; Gaps 4;
QY 7 SPRVGVFVAGRMAGATAOGLIRAGKVEAQHILASAPTDRLNCHFGALGCRTHSNQEV 66
Db 4 SMARIAIIGGSGTEALLSGLLRAGROVKDLVVAERPDRARYLVLT-SYTDV 62
QY 67 QSCLLVIFATKPHVLPVLAIEVAPVW-----TTEHILVSVAGVSLSTLEELLPNTRVL 121
Db 63 ENAMFVVVAVKPTDVESVMDGLVQAAVAANDSAEQVLVTVAAGVTITYLESKLPAGTPV 122
QY 122 RVLNPLFCVQEGAIVMARGHVGSSETKLQHLLEACGRCCEVPEYVDIHTGLSGSV 181
Db 123 RAMENAAALVGAGVTVLAKGRFVTGQOFEDVLFAMFVAGVGLTVLPESQMDAVTAVSGSGP 182
QY 182 AFVCAFSBALAEGAVKMGMPSSLAHRTAAOTLLGTAKMLLH---EQH----- 226
```

```
Db 183 AYFELLVEALVDAGVAVGLTRQVATELAAQTMAAGSAAMLLRMDQDRHSNAEVLPGAQVD 242
QY 227 --PAQLRSDVCTPGGTTIYGLHLEOGLGRAATMSAVEAATCAKEL 271
    ||||| : ||||| | ||||| : ||||| : ||||| : |||||
Db 243 VPAQLRATITSPGGTTAAALRELGRGLRMVDDVAQAQAKIRSEOL 289

RESULT 11
PROH_BACSU
ID PROH_BACSU STANDARD; PRT; 297 AA.
AC P14383; O31928; O07508;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyroline-5-carboxylate reductase 1 (EC 1.5.1.2) (P5CR 1) (P5C
DE reductase 1).
GN PROH OR PROC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Hega K., Hache J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone E.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yanane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL [4]
RN [2]
RP SEQUENCE OF 1-256 FROM N.A.
RC STRAIN=W23, and 168;
RX MEDLINE=91192601; PubMed=1849493;
RA Ahn K.S., Wake R.G.;
RT "Variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W23 chromosomes."
RL Gene 98:107-112(1991).
RN [3]
RP SEQUENCE OF 1-200 FROM N.A.
RC STRAIN=W23;
RX MEDLINE=89155440; PubMed=2493444;
RA Lewis P.J., Wake R.G.;
RT "DNA and protein sequence conservation at the replication terminus in
RT Bacillus subtilis 168 and W23."
RL J. Bacteriol. 171:1402-1408(1989).
RN [4]
RP SEQUENCE OF 199-297 FROM N.A., AND FUNCTION.

RC STRAIN=168 / SMY;
RX MEDLINE=21311767; PubMed=11418582;
RT Belitsky B.R., Brill J., Bremer E., Sonenshein A.L.;
RT "Multiple genes for the last step of proline biosynthesis in Bacillus
RT subtilis."
RL J. Bacteriol. 183:4389-4392(2001).
CC -!- CATALYTIC ACTIVITY: L-proline + NAD(P)(+) -> L-pyrroline-5-
CC carboxylate + NAD(P)H.
CC -!- PATHWAY: Proline biosynthesis; third (last) step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PYRROLINE-5-CARBOXYLATE REDUCTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z991114; CAB13741.1; -.
DR EMBL; W24523; AAA22722.1; -.
DR PIR; B32807; B32807.
DR Subtilist; BG11049; proH.
DR InterPro; IPR000304; P5CR.
DR Pfam; PF01089; P5CR; 1.
DR TIGRFAMS; TIGR00112; proC; 1.
DR PROSITE; PS00521; P5CR; 1.
KW Oxidoreductase; Proline biosynthesis; NADP: Complete proteome.
FT VARIANT 58 58 T -> A (IN STRAIN W23).
FT VARIANT 70 70 L -> S (IN STRAIN W23).
FT VARIANT 74 74 L -> I (IN STRAIN W23).
FT VARIANT 93 93 N -> S (IN STRAIN W23).
FT VARIANT 100 100 S -> T (IN STRAIN W23).
FT VARIANT 127 127 E -> Q (IN STRAIN W23).
FT VARIANT 159 159 K -> Q (IN STRAIN W23).
FT VARIANT 236 236 H -> Q (IN STRAIN W23).
FT VARIANT 239 239 I -> V (IN STRAIN W23).
FT CONFLICT 256 256 Q -> E (IN REF. 2).
FT CONFLICT 264 267 GFALSAIKHAARKSEISIEDIEKTAAPLSGVIK -> EKR
SQ SEQUENCE 297 AA; 32031 MW; F4DC7B6F84C6ED42 CRC64;
Query Match 27.3%; Score 378; DB 1; Length 297;
Best Local Similarity 32.2%; Pred. No. 2.4e-22;
Matches 91; Conservative 50; Mismatches 106; Indels 36; Gaps 3;
QY 9 RVGVFVAGRMAGTAQGLIRAGKVEAQHILASAPTRNLCHFQALGCRTHSNOEVL-- 66
Db 17 KKVAFVAGSMAGMISGIVRANKIPKONI-----CVTNKSNTERLUTE 59
QY 67 -----QSCL-----LVIPATKPHVLPVLAEPVVTTEHILSVAAAGVSLST 109
Db 60 LELOYGKALPNQCIEDMDVLLAMKPKDAENALSLKSIQPHQLLSVLACITTSF 119
QY 110 LEELPPNTRVLRVLPNLPVVOEGATVNRGHRVGSSETKLLQLHLBACGCEVPBAY 169
Db 120 IEQSLINEQPVVVRVMPNTSSMIGASATAIALGKYVSEDLKLAELALGCMGEVYTIQENQ 179
QY 170 VDIHTGLSGSVAFVFCAPSEALAEAGVKMGPMSSLAHRTAAOTLLGTAKMLLHCOHPAQ 229
Db 180 MDIFITAGSGPAYFYLLMEFTEKTGEEAGLDKQLSRISGIAOTLLGAARKMLMETGEHPEI 239
QY 230 LRSDVCTPGGTTIYGLHLEOGLGRAATMSAVEAATCAKEL 272
Db 240 LRDNITSPNGTTAAGLQALKSGGGEATSOAIKHAARKSEIS 282
RESULT 12
PROC_MYCTU
ID PROC_MYCTU STANDARD; PRT; 295 AA.
```


